(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 March 2001 (29.03.2001)

PCT

(10) International Publication Number
WO 01/21839 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US00/26055

(22) International Filing Date:
22 September 2000 (22.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/156,050 24 September 1999 (24.09.1999) US
60/209,026 2 June 2000 (02.06.2000) US

(71) Applicant *(for all designated States except US)*: PHARMACIA & UPJOHN COMPANY [US/US]; 301 Henrietta Street, Kalamazoo, MI 49001 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: FOSTEL, Jennifer, M. [US/US]; 5047 Angling Road, Portage, MI 49024 (US). BAMMERT, Gary, F. [US/US]; 5339 Woodmont Drive, Portage, MI 49002 (US).

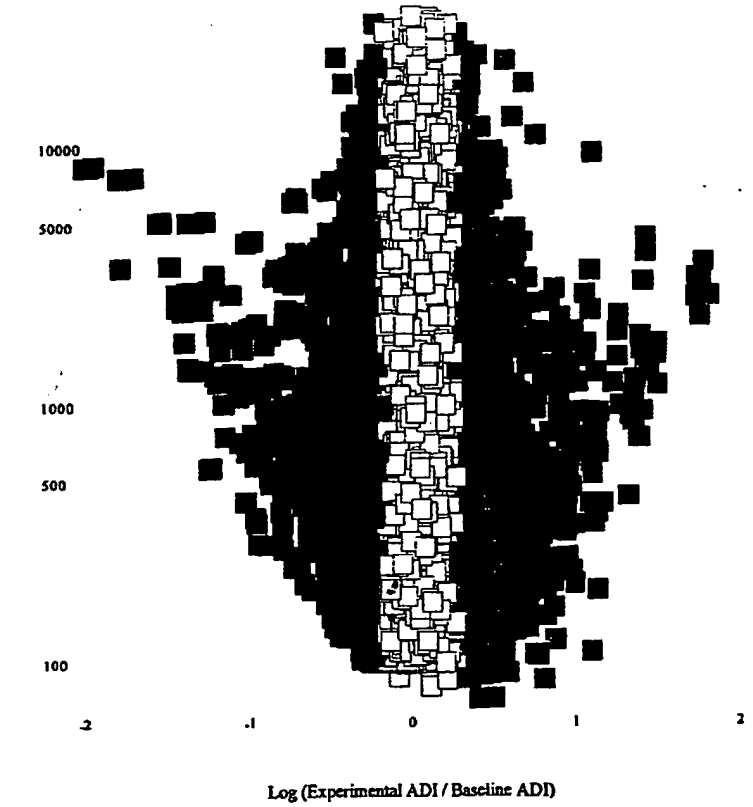
(74) Agents: MILLER, Suzanne, E. et al.; Woodcock Washburn Kurtz Mackiewicz & Norris LLP, 46th floor, One Liberty Place, Philadelphia, PA 19103 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

[Continued on next page]

(54) Title: GENE EXPRESSION PROFILES



(57) Abstract: This invention provides a simplified method of analysis of data generated by microarrays permitting those of skill in the art to generate expression profiles of pathogens. The methods of the invention allow the rapid screening of candidate drug compounds, identification of drug candidates predicted to be effective against particular pathogens, and identification of genes involved in the pathway affected by such drug compounds.

WO 01/21839 A2

patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

GENE EXPRESSION PROFILES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Patent Applications, Serial Number 60/156,050 filed September 24, 1999 and Serial Number 60/209,026 filed June 2, 2000, which are hereby incorporated by reference.

FIELD OF THE INVENTION

This invention is related to the fields of gene expression patterns, of identifying agents with antibiotic function, and of identifying genes responsive to such agents.

BACKGROUND OF THE INVENTION

Treatment of infections with antibiotics has led to the development by many pathogens of clinical resistance to such antibiotics. In the case of fungal pathogens, opportunistic fungal infections have become a life-threatening problem for individuals with compromised immune systems, and azoles represent a significant portion of drugs used to treat systemic infection by fungal commensal organisms. For example, fluconazole has emerged as the primary therapy for the treatment of oropharyngeal candidiasis in HIV-infected patients (Revankar *et al.*, 1996, J. Infect. Dis. 174:821-827; Rex *et al.*, 1995, Antimicrob. Agents Chemother. 39:1-8). Azoles such as fluconazole inhibit lanosterol 14a-demethylase (Erg11p), an enzyme in the ergosterol biosynthetic pathway in yeast (Turi & Loper, 1992, J. Biol. Chem. 267:2046-2056). The nitrogens in the azole ring form a complex with the heme iron component of the cytochrome group resulting in the inhibition of the enzyme (Yoshida & Aoyama, 1987, Biochem. Pharmacol.

36:229-235).

Widespread treatment with azoles has led to clinical resistance of many *Candida* spp. and other fungal pathogens (Denning *et al.*, 1997, Antimicrob. Agents Chemother. 41:1364-1368; Joseph-Horne *et al.*, 1995, Antimicrob. Agents Chemother. 39:1526-1529; Moran *et al.*, 1998, Antimicrob. Agents Chemother. 42:1819-1830; Vanden Bossche *et al.*, 1992, Antimicrob. Agents Chemother. 36:2602-2610; Venkateswarlu *et al.*, 1996, Antimicrob. Agents Chemother. 40:2443-2446). Isolation of these resistant strains has led to the intensive study of the molecular mechanisms by which the organism can compensate for growth in the presence of these drugs. Clinical isolates of azole-resistant *C. albicans* have been shown to contain point mutations in the gene encoding Erg11p (Lamb *et al.*, 1997, J. Biol. Chem. 272:5682-5688; White, 1997, Antimicrob. Agents Chemother. 41:1488-1494) or an increased expression of this gene (White, 1997, Antimicrob. Agents Chemother. 41:1482-1487) leading to decreased susceptibility to azole drugs. In addition to altering the target gene, *C. albicans* has been shown to increase the efflux of azoles by upregulation of multidrug resistant pumps belonging to the ABC transporter and major facilitator families in response to azole exposure, resulting in lower susceptibility of the organism (Sanglard *et al.*, 1995, Antimicrob. Agents Chemother. 9:2378-2386; White, 1997, Antimicrob. Agents Chemother. 41:1482-1487.).

Once the genome of a pathogen has been partially or fully sequenced, genechip microarrays may be constructed containing open reading frames (ORFs) spanning the entire genome. For example, following the complete sequence determination of the *Saccharomyces cerevisiae* genome, DNA microarrays have emerged as a powerful tool for examining the simultaneous expression pattern of more than 6000 yeast genes (DeRisi *et al.*, 1997, Science. 278:680-686; Gray *et al.*, 1998, Science. 1281:523-528; Jelinsky & Samson, 1999, Proc. Natl. Acad. Sci. USA. 96:1486-1491; Marton *et al.*, 1998, Nature Medicine 4:1293-1301; Wodicka *et al.*, 1997, Nature Biotech. 15:1359-1367). However, analysis of the data resulting from the use of such microarrays has heretofore been plagued by the need for computation of complex algorithms which limits the speed at which one of skill in the art may use such data to identify useful information such as drug candidates and target genes.

For example, Tavazoie *et al.* (1998, Nature Genetics 22:281-285) used the iterative *k*-means algorithm to identify transcriptional regulatory sub-networks in yeast, from hybridization data generated from 15 time points, spanning two cell cycles of growth. This entailed a highly complex and time-consuming analysis of the hybridization intensities resulting from microarray screening.

Alternatively, in Gray *et al.* (*supra*), microarray screening was used to identify transcripts in yeast whose levels were altered more than two-fold when treated with kinase-inhibitory purine derivatives, without subjecting the data to any further analyses.

In Wodicka *et al.* (*supra*), genome-wide expression levels were compared for yeast cells grown in rich and minimal media. RNA level or abundance was determined by a quantitative analysis of the hybridization intensities on hybridized microchips according to the method of Lockhart *et al.* (1996, Nature Biotech. 14:1675-1680). Yet, this form of analysis does not allow one to easily distinguish treatment-specific responsive genes.

Marton, *et al.* (*supra*) examined the genome-wide effect of gene deletions and chemical inhibitors of the calcineurin-signaling pathway on gene expression in yeast. Comparisons between two treatments, or between a treatment and control, were made by plotting the log of the expression ratio (control culture/treated culture) for each ORF for one treatment or mutation, against the log of the expression ratio for each ORF for another treatment or mutation, or against the log of the mean intensity of hybridization for each ORF.

The art is in need of a simplified method of analyzing microarray data, in order to speed the identification of drug candidates, target genes, and information related to modes of action of such drugs and steps in biosynthetic pathways.

SUMMARY OF THE INVENTION

This invention provides a simplified method of analysis of data generated by microarrays permitting those of skill in the art to generate expression profiles of agent-treated pathogens. The methods of the invention allow the rapid screening of candidate drug compounds, identification of drug candidates predicted to be effective against particular pathogens, and identification of genes involved in the pathway affected by such

drug compounds.

For example, the present invention is directed to a method of preparing an expression profile of genes expressed by a cell treated with an agent, comprising the step of plotting the log(ratio) of normalized filtered ADI values resulting from a hybridization assay performed on a microarray with mRNA extracted from the cell against a sum of the normalized filtered ADI values. The method may be carried out with an antibiotic agent or an agent which is not an antibiotic agent. Expression profiles prepared by this method are also part of the invention.

The invention is also directed to a method of identifying an agent capable of altering expression of a gene in a cell, comprising the step of identifying a treatment-affected gene from a prepared expression profile, wherein the treatment-affected gene's datapoint on the expression profile exceeds one standard deviation from a distribution, as well as agents identified thereby. Preferably the datapoint exceeds two standard deviations from the distribution.

The invention is also directed to a method of identifying a first agent capable of interfering with a biosynthetic pathway, comprising the steps of comparing a plurality of first treatment-affected genes identified from an expression profile of a first cell treated with the first agent, with a plurality of second treatment-affected genes identified from an expression profile of a second cell treated with a second agent known to interfere with the biosynthetic pathway; and determining whether the number of the first treatment-affected genes which are identical to the second treatment-affected genes exceeds a percentage of about 20% of the total number of second treatment-affected genes, as well as agents identified thereby.

The invention is also directed to a method of identifying a gene as part of a biosynthetic pathway comprising the step of identifying a treatment-affected gene common to a percentage of expression profiles prepared from cells treated with agents known to affect the biosynthetic pathway, as well as genes, in isolated form, identified thereby.

These and other aspects of the invention are described in greater detail below.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the growth curves of *S. cerevisiae* determined using the Spectromax Plus plate reader (diamonds) and shake flasks (circles).

Figure 2 shows the determination of the minimum inhibitory concentration (MIC) of ketoconazole for *S. cerevisiae* strain BY4743. Cultures were grown for 24 hours and OD_{600} measured in the Spectromax Plus plate reader.

Figure 3 depicts the relationship between the treatments used and ergosterol biosynthesis. Gene names were as listed in Daum *et al.*, 1998, Yeast 14:1471-1510.

Figure 4 shows the expression profiles obtained from this study. Each point represents the response of a gene to a particular treatment. ORFs with Average Difference Intensity (ADI) values below 50 ADI units were not included. Plotted is the sum of experimental ADI and baseline ADI for each gene (vertical axis), and, on the horizontal axis,

$$\log(ratio)_i = \log((E_i / BL_i) / \sqrt[n]{\prod_{j=1} E_j / BL_j})$$

Figure 5 shows the genes involved in the biosynthesis of ergosterol and membrane components. Genes given in bold type were responsive in the study; italicized bold type indicates genes with decreased transcript level. Superscripts indicate the number of treatments to which the gene responded. The first number in the superscript indicates the number of the three genetic perturbations which elicited a response, and the second number indicates the number of the eight drug treatments which elicited a response. Lists of genes were obtained from the Yeast Protein Database (YPD) (Hodges *et al.*, 1999, Nuc. Acids Res. 27:69-73) and Daum *et al.* (*supra*) and Denning *et al.* (*supra*).

Figure 6 shows the genes involved in biosynthesis and utilization of heme. Bold type and superscripts are as described for Figure 5. Lists of genes were obtained from the YPD database (Hodges *et al.*, *supra*).

Figure 7 shows the PCR measurements of transcripts from five genes: ERG3, ERG4, ERG5, ERG6, ERG11/16, and ERG24. Shown is the change with respect to the internal control, TEF1. An increase in cycle number indicates decrease in initial template

concentration.

Figure 8 presents comparisons of qualitative gene expressions patterns in response to different biological treatments. 18 functional pathways are presented along the x-axis. The treatments are presented along the y-axis.

DETAILED DESCRIPTION OF THE INVENTION

It will be appreciated by those of skill in the art that the methods of the present invention are useful for analysis of many different pathogens, including but not limited to bacteria, yeast, and viruses, as well as eukaryotic cells. It is preferred that the pathogen's entire genome be sequenced, however, partial sequencing still permits the formulation of gene expression profiles and would thus be appropriate for the methods of the present invention.

In general, the methods of the invention entail exposing growing cells to agents of interest, and analyzing the mRNA transcripts present in the cells, in response to the exposure, preferably by hybridization to DNA microarrays. Cells are exposed to agents and harvested preferably during mid-logarithmic growth phase. mRNA is isolated and typically converted to double stranded cDNA, so that amplified and labeled cRNA may be generated. The labeled cRNA is then incubated with a DNA microarray to allow for hybridization of complementary sequences. After unbound cRNA is removed, the hybridized material is detected. Detection of hybridized cRNA may be accomplished by any known method, preferably by binding of streptavidin-phychoerythrin to the biotin-labeled cRNA, and scanning for the phychoerythrin signal. Alternatively, the cRNA itself is labeled with fluorescently-tagged nucleotides, during amplification, for direct detection. The signals are manipulated by software (in the case of using an Affymetrix microarray, by the Affymetrix Genechip algorithm (using Affymetrix software)) to yield hybridization intensity values (average difference intensity or ADI units) for each gene on the array.

Affymetrix microarrays consist of oligonucleotides synthesized *in situ*. Each ORF has a number of corresponding oligomers on the microarray, termed a probe set. A probe set consists of a series of probe pairs: oligomers designed to match the ORF sequence ("perfect match" or PM), and, for each, a corresponding mismatch (MM) oligomer

designed to serve as a hybridization control. The Affymetrix Genechip algorithm computes the hybridization signal, termed Average Difference Intensity (ADI), for the probe pairs whose difference lies within three standard deviations of the differences of other probe pairs from the same probe set. This exclusion of outliers is done to minimize false signals due to bright specks or non-responding probe pairs. The ADI value for each ORF is the sum of (PM-MM) differences for the probe pairs used, divided by the number of probe pairs used in the calculation.

To obtain ADI values above the chip background, ADI values from the experimental data sets less than or equal to 50 are eliminated (filter A). A second filter is then applied requiring baseline ADI to be greater than 50 (filter B).

Prior to analysis, ADI values are normalized, i.e., edited to replace values of 1 or below with a value of 1. These include negative values that arise from ORFs in which the "Mismatch" signal is greater than the "Perfect match" signal, and reflect non-expressed ORFs. The actual value of these negative ADI differences is not pertinent to the analysis of expressed genes, and replacing these values with 1 eliminates subsequent editing required to remove errors generated when attempting to divide by 0 or to calculate the logarithm of a negative number. Following this normalization procedure, the Experimental (E) ADI values are compared on a gene-by-gene basis to the ADI value of the ORF under unperturbed conditions (termed the Baseline (BL) value, the arithmetic average of ADI values for that ORF derived from six replicate data sets from mid-log phase growth). Two values are computed for each ORF: (1) the sum of edited ADI values for that ORF from Experimental and Baseline data sets, and (2) the log of the ratio (log(ratio)) of Experimental to Baseline ADI values for that ORF.

To compensate for differences between chips in a data set, the E/BL ratio for each ORF is divided by the geometric mean of the E/BL ratios for that chip. Dividing by the geometric mean has the effect of centering the log(ratio) distribution on zero and eliminating any small variance between chips within a data set prior to genome-wide analysis.

Thus the values (composed of Experimental (E_i) ADI and BaseLine (BL_i) ADI values) used for analysis are, for each $gene_i$ of the n genes on the chip:

$$sum_i = E_i + BL_i$$

$$\log(ratio)_i = \log((E_i / BL_i) / \sqrt[n]{\prod_{j=1}^{n} E_j / BL_j})$$

Comparisons between data sets are preferably performed visually using Spotfire Pro 3.0 (Spotfire Corp., Cambridge MA) and numerically using Microsoft Excel. Data are imported into Spotfire, binned by standard deviation, and displayed graphically. The graphical representation of this data, an example of which is shown in Figure 4, represents a novel method of formulating an expression profile, and is useful for all the methods of the invention.

In one embodiment of the invention, the log(ratio) (as defined above) for each ORF is plotted against the sum (as defined above) for each ORF, yielding an expression profile such as that shown in Figure 4. As a control, in a preferred embodiment, such profiles may be prepared from untreated cells as well as from cells treated with potential active agents, and the profiles compared.

In a preferred embodiment, the datapoints on the expression profile are considered in light of their standard deviations from the distribution, as is also shown in Figure 4. Preferably those ORFs showing greater than one standard deviation from the distribution in either the positive or negative direction are considered indicative of up- or down-regulated genes, respectively, and are thus identified as treatment-affected genes. More preferably, those ORFs showing greater than two standard deviations from the distribution in either the positive or negative direction are considered indicative of up- or down-regulated genes, respectively (i.e., treatment-affected genes).

In another embodiment, the expression profiles are used to evaluate potential active agents which may or may not belong to a class of agents known to interfere with a particular biosynthetic pathway. Cells are treated with agents known to interfere with the biosynthetic pathway, and expression profiles are created as in previously discussed embodiments. The expression profiles are compared in order to identify those genes which are up- or down-regulated by the active agents. In a preferred embodiment, those genes whose expression is most affected by the active agent (e.g., show greater standard deviations from the distribution) are identified as treatment-affected genes. Preferably, the

most affected genes show standard deviations in excess of one standard deviation, more preferably in excess of two standard deviations.

Expression profiles from cells treated with agents whose effect is unknown may then be prepared and compared to the expression profiles for active agents known to interfere with biosynthetic pathways, in order to determine whether the agent whose function is unknown would be predicted to have the same or similar effects as the known agents. In a preferred embodiment, an unknown agent whose expression profile shows up- and down-regulation of some of the treatment-affected genes allows one of skill in the art to conclude that the unknown agent is likely to be effective at interfering with the biosynthetic pathway in question. Preferably, the agent whose function is unknown can be said to be likely to have similar effects as the known-active agents if the treatment-affected genes identified in the expression profile of the unknown agent are some of the same treatment-affected genes identified in the expression profiles of the known agents. Preferably, if about 20% of the treatment-affected genes identified in the expression profile of the known active agents also are identified as treatment-affected genes in the expression profile of the unknown agent, it can be concluded that the unknown agent is likely to exert substantially similar effects on the pathway involved as the known agents. More preferably, the percentage of treatment-affected genes common to the expression profiles of both the known and unknown agents allowing one of skill in the art to conclude a substantially similar effect on the pathway is about 30%, more preferably about 40%, more preferably about 50%, more preferably about 60%, more preferably about 70%, more preferably about 80%, more preferably about 90%, and most preferably about 100%.

In another embodiment of the invention, the expression profiles prepared as described above are used to identify the genes affected by a panel or class of agents. Several members of a class of agents with known activity against a particular pathogen are used to treat target cells, and expression profiles are prepared as described above. The treatment-affected genes are then identified for each expression profile. Preferably, the treatment-affected genes are identified as having standard deviations in excess of one standard deviation, more preferably in excess of two standard deviations.

In this document agents are considered any composition of matter, preferably small

molecules, that can do the following: they might inhibit the growth of a target cell, they might prove to be biologically active either enhancing or ameliorating a particular disease state. Agents can also produce nonspecific effects, such as producing general physiological stress. As used here, for example, an antibiotic is an agent that inhibits a microorganism. The microorganism may be a eukaryotic or prokaryotic cell, it can include yeast and fungus, bacteria, protozoa. Here when an antifungal agent is referred to it would be an antibiotic agent that inhibits the growth of yeast or fungi.

In a preferred embodiment, the treatment-affected genes identified for the expression profiles of each active agent are compared, and treatment-affected genes identified as common to at least about 25% of the class of active agents tested are identified as class-specific treatment-affected genes. Preferably, treatment-affected genes identified as common to at least about 40% of the class of active agents tested are identified as class-specific treatment-affected genes. Preferably, treatment-affected genes identified as common to at least about 55% of the class of active agents tested are identified as class-specific treatment-affected genes. Preferably, treatment-affected genes identified as common to at least about 70% of the class of active agents tested are identified as class-specific treatment-affected genes. Preferably, treatment-affected genes identified as common to at least about 85% of the class of active agents tested are identified as class-specific treatment-affected genes. Preferably, treatment-affected genes identified as common to at least about 100% of the class of active agents tested are identified as class-specific treatment-affected genes.

In a preferred embodiment, the class-specific treatment-affected genes are considered regarding their role in a biosynthetic pathway. Where a gene or several genes are involved in the same pathway, it may be concluded that it is likely that the class of active agents exert their effects on that particular pathway. This method thus also aids in identification of the mode of action of a class of agents.

In another embodiment, the invention is directed to the expression profiles of yeast treated with the azole and imidazole compounds described above.

Other embodiments of the invention will be readily understood by those of skill in the art.

The present invention is more fully described using the particular example of yeast. although the methods used for other organisms would be essentially the same.

Ergosterol is an essential component of yeast plasma membranes which affects membrane fluidity, permeability, and the activity of membrane bound enzymes (Cobon & Haslam, 1973, Biochem. Biophys. Res. Comm. 52:320-326; Daum *et al.*, *supra*; Parks, 1987, CRC Crit. Rev. Microbiol. 6:301-341). In *S. cerevisiae*, ergosterol is also a major component of secretory vesicles and has an important role in mitochondrial respiration (Daum *et al.*, *supra*; Parks, *supra*; Zinser *et al.*, 1993, J. Bact. 175:2853-2858). Ergosterol has also been predicted to play a role in oxygen sensing (Smith & Parks, 1997, Biochem. Biophys Acta 1345:71-76), defined by the well-characterized sparking function of this sterol (Lorenz *et al.*, 1989, J. Bact. 171:6169-6173; Rodriguez *et al.*, 1985, Biochem. Biophys. Acta 837:336-343). Genes in the ergosterol pathway demonstrate transcriptional regulation in response to a mutation in other ERG genes and resulting sterol limitation (Arthington-Skaggs *et al.*, 1996, FEBS Lett. 392:161-165; Kennedy *et al.*, 1999, Biochim. Biophys. Acta. 1445:110-122). Other genes such as CYB5, COX3, RPL27 among others have been shown to contribute to altered sensitivity to azoles when overexpressed (Launhardt *et al.*, 1998, Yeast, 14:935-942; Truan *et al.*, 1994, Gene. 142:123-127). Thus, the analysis of the genome-wide response to ergosterol perturbation may reveal novel mechanisms for resistance, or new sites for chemical intervention, in addition to increasing understanding of the cellular response to perturbation of ergosterol biosynthesis.

The methods of this invention were used to reveal a convergent pattern of gene expression between drug-treated cells and cells with genetic alterations in the same targeted pathway. From the data generated, a set of genes can be identified that is predictive of an "ergosterol response." Other responsive genes have been identified that offer insight into the relation of ergosterol biosynthesis to important physiological changes in the cell. Additionally, informative transcriptional changes were observed that can be used to identify potential targets for chemical intervention. The invention also provides methods by which genome-wide transcriptional changes in yeast can be analyzed following exposure to a drug with an uncharacterized mode of action.

Inhibitors of ergosterol biosynthesis such as the azoles are a major class of antifungal therapy. Disruption of ergosterol biosynthetic genes mimics phenotypic changes observed during exposure to these agents, thus lending evidence to the agents' mode of action. Using Affymetrix microarrays, gene expression patterns were obtained for *S. cerevisiae* strains bearing homozygous disruptions of ERG2, ERG5 or ERG6 and their parent strain BY4743. In an effort to demonstrate a correlation between a genetic mutation and a biological drug response, the transcript profile was determined following exposure of BY4743 cells to sub-MIC levels of a number of inhibitors of ergosterol biosynthesis. Comparison of these transcript profiles lends insight into known and previously unknown modulations of gene expression. Genes involved with general cellular response to these agents can be distinguished from genes modulated in cells carrying a single genetic aberration in this pathway. A set of marker genes was ascertained from these transcript profiles, and primer/probe sets designed to verify these transcription patterns using quantitative RT-PCR. This method confirms the changes in transcript levels observed using hybridization to microarrays.

Enzymes in the yeast ergosterol biosynthetic pathway are the targets of a number of antifungal agents including azoles, allylamines, and morpholines. In order to understand the response of *S. cerevisiae* cells to perturbations in the ergosterol pathway, genome-wide transcript profiles following exposure to a number of antifungal agents targeting ergosterol biosynthesis (clotrimazole, fluconazole, itraconazole, ketoconazole, voriconazole, terbinafine and amorolfine) were obtained. These profiles were compared to the transcript profiles of strains containing deletions of late-stage ergosterol genes: ERG2, ERG5, or ERG6. This comparison allowed the response to chronic genetic perturbations to be compared to the response resulting from acute exposure to chemical agents. Data sets from these experiments were compared on a gene-by-gene basis and filtered to derive a subset of genes exhibiting a transcriptional response to ergosterol perturbation (i.e., class-specific, treatment-affected genes). Among the 234 genes identified, the majority of genes from the ergosterol pathway were found to be responsive. The levels of transcripts from ERG3, ERG4, ERG5, ERG6, ERG11 and ERG24 were measured over time by quantitative PCR. Comparison of these measurements to microarray results showed

similar trends in expression levels. In addition to ergosterol biosynthetic genes, 36 mitochondrial genes and a number of other genes with roles related to ergosterol function were responsive, as were a number of genes responsive to oxidative stress. Transcriptional changes related to heme biosynthesis were observed in cells treated with chemical agents, suggesting an additional effect of exposure to these compounds. The expression profile in response to a novel imidazole, PNU-144248E, was also determined. The concordance of responsive genes suggests that this compound has the same mode of action as other imidazoles. Thus, genome-wide transcript profiles can be used to predict mode of action as well as to characterize expression changes in response to perturbing a metabolic pathway.

The invention is further illustrated by way of the following examples, which are intended to elaborate several embodiments of the invention. These examples are not intended, nor are they to be construed, as limiting the scope of the invention. It will be clear that the invention may be practiced otherwise than as particularly described herein. Numerous modifications and variations of the present invention are possible in view of the teachings herein and, therefore, are within the scope of the invention.

EXAMPLES

Example 1: Materials and Methods for Practicing the Invention with Yeast

Reagents

Clotrimazole, biotin-11-CTP and biotin-16-UTP were obtained from Sigma Chemical Company; ketoconazole was from Biomol Research Labs, Inc.; itraconazole was from Accurate Chemical and Scientific Corp., Westbury NY; and terbinafine was from TCI, Tokyo. Amorolfine, fluconazole, voriconazole and PNU-144248E were synthesized at Pharmacia & Upjohn.

Yeast Strains

S. cerevisiae BY4743 (a/a his3Δ/his3Δ, leu2Δ/leu2Δ, +/lys2Δ, met15Δ/+, ura3Δ/ura3Δ), strain 30568 (a/a, erg6Δ/erg6Δ, his3Δ/his3Δ, leu2Δ/leu2Δ, +/lys2Δ, met15Δ/+, ura3Δ/ura3Δ), strain 30590 (a/a, erg5Δ/erg5Δ, his3Δ/his3Δ, leu2Δ/leu2Δ, +/lys2Δ, met15Δ/+, ura3Δ/ura3Δ), strain 30788 (a/a, erg2Δ/erg2Δ, his3Δ/his3Δ,

leu2Δ/leu2Δ, +/lys2Δ, met15Δ/+, ura3Δ/ura3Δ). All strains were obtained from Research Genetics (http://www.resgen.com).

MIC Determinations

A 10 ml culture of YPD medium (1% Difco yeast extract, 2% Difco peptone, 2% glucose) was inoculated from a colony and grown overnight at 30° C to saturation (~ 1 to 2 x 10^8 cells/ml). The culture was then diluted to an OD_{600} = 0.1 and 50 ml used to inoculate a 96 well U-bottom culture plate (Costar, Corning, NY) containing 50 ml of 2-fold serially diluted test compounds starting at 100 mM. The culture was allowed to grow in the presence of drug for 24 hrs at 30°C in a SpectraMax Plus (Molecular Dynamics Corp., Sunny Vale, CA) and the minimum inhibitory concentration (MIC) was determined as the first well with no growth. The growth curve obtained for each drug was used to identify a sub-MIC concentration needed to slow growth without affecting the endpoint cell concentration. The endpoint cell concentration was required to be sufficiently high enough to yield an adequate amount of poly A(+) RNA for expression analysis.

Cell Culture and Drug Exposure

A 40 ml culture of YPD medium was inoculated from a colony and grown overnight at 30° C and 140 rpm to saturation (~ 1 to 2 x 10^8 cells/ml). The culture was diluted ten-fold to allow growth through one doubling (90 min). For drug-treated cells, drug was added to each culture at a concentration equal to 0.5 X MIC (concentrations used were: 0.6 μM clotrimazole, 25 μM fluconazole, 0.6 μM itraconazole, 4 μM ketoconazole, 0.6 μM PNU-144248E, 0.19 μM voriconazole, 0.1 μM amorolfine, and 50 μM terbinafine), and the culture incubated for 90 minutes. Strains containing mutations were harvested at late-logarithmic phase (~ 5 x 10^7 cells/ml). Cells were pelleted at 1500 rpm 20° C for 5 min in an IEC MP4R centrifuge. The pellets were washed with 1 ml water at 22°C, previously treated with diethylpyrocarbonate (DEPC) to inactivate RNases. Cell pellets were placed on ice and RNA was extracted immediately to minimize change in expression profile.

RNA Preparation and Hybridization to Affymetrix Microarrays

RNA preparation and hybridization to Affymetrix DNA microarrays were performed as described by Wodicka *et al.*, *supra*. Briefly, cells were harvested, washed

with water and lysed quickly and RNA extracted using hot acidic phenol (Lin *et al.*, 1996, pp.43-50, in P.A. Kreig (ed.), A laboratory guide to RNA isolation and synthesis. Wiley-Liss, Inc., New York). Poly A(+) RNA isolated using Oligotex mRNA kit from Qiagen Inc. (Valencia, CA). Double-stranded cDNA synthesized using the Superscript Choice system (GibcoBRL, Gaithersburg, MD). Labeled cRNA was generated with biotin-11-CTP and biotin-16-UTP using the Megascript System (Ambion Inc., Austin TX) for hybridization to the microarrays. Eleven mg of the resulting cRNA was used to probe the four arrays comprising the yeast genome, following the method recommended by Affymetrix Inc. (Santa Clara, CA) and described by Wodicka *et al.*, *supra*.

Experimental Design and Methodology

Because of the scope of each data set obtained from a microarray hybridization, exposures were limited to a single concentration at a single time point. The concentration and duration of exposure were chosen to elicit a biological effect while attempting to minimize secondary effects due to prolonged exposure. Since *S. cerevisiae* will double approximately every 90 minutes in rich media (as shown in Figure 1), one doubling time was selected for the duration of exposure. Sub-MIC concentrations sufficient to slow but not stop growth were used, as described above and as illustrated in Figure 2. Cells were exposed to agents and harvested during mid-logarithmic growth phase when the cell is most rapidly dividing and expression patterns seen in slow-growing or saturated cultures are not observed. The *S. cerevisiae* strain used in this study was BY4743, the parent strain for homozygous deletion strains used, and isogenic to that used for sequence determination and microarray synthesis.

RNA was prepared following exposure of the parent strain BY4743 to five previously characterized azoles (clotrimazole, fluconazole, itraconazole, ketoconazole, voriconazole), a novel imidazole, PNU-144248E, (as shown in Figure 3), an allylamine (terbinafine) and a morpholine (amorolfine) as described above. In addition, RNA was prepared from three strains each bearing a homozygous deletion of ERG2, ERG5, or ERG6 encoding C-8 sterol isomerase, sterol C-22 desaturase, and sterol C-26-methyltransferase respectively. RNA preparations were also made from two untreated control cultures.

Figure 3 also shows the genes involved in ergosterol biosynthesis from farnesyl pyrophosphate in *S. cerevisiae*. Genes in the deletion strains are given in bold font. The site of action of the azole, allylamine, and morpholine antifungal agents are noted with arrows, and the structures of the compounds used in this study are given.

Example 2: Identification of Gene Expression Patterns

Each Affymetrix yeast genome set represents 6593 ORFs, including 172 control genes. Thus the eleven data sets from the treatments comprising this study represent 72523 data points. Hybridization intensity values are expressed in ADI units, as described above. Following normalization to account for variations in chip intensity (described above), filter "A" was applied requiring the experimental ADI value to be above 50 ADI units, thereby limiting the data to 29428 points. Applying a second filter, "B", requiring the baseline ADI value to be greater than 50 ADI units, further limits the data to 20697 points. The ADI values used in the filters were chosen after examination of the background ADI value calculated by the GeneChip software and the ADI values for selected unexpressed genes (e.g. haploid-specific genes in diploid cells); both approaches gave a value of up to ~25 ADI units as that corresponding to no signal.

Experimental ADI values were then expressed on a gene by gene basis as the sum of experimental ADI with baseline ADI, and as the $\log_{10}$ of the ratio of the experimental ADI to the baseline ADI. To account for inter-chip variation in signal intensity, log(ratio) distributions obtained for each chip were centered by subtracting the geometric mean from the individual values as described above. This has the effect of multiplying ADI values by at most a factor of three in the experiments described here, and allows comparison of hybridization values obtained from different chips. These values are plotted in Figure 4. Points are colored by position in the distribution of log ratios, using white for points within one standard deviation of the distribution, black for those points between one and two standard deviations of the distribution, and gray for those beyond two standard deviations.

Genes with ADI ratios beyond one standard deviation were considered to be responsive to the treatment (i.e., treatment-affected genes). By this criterion, 1154 genes

responded by increased mRNA level in at least one treatment; 1358 genes responded by decreased levels relative to baseline. It is likely that expression changes specific for the particular treatment as well as other perturbations in transcript level due to uncontrolled changes in microenvironment, unstable transcripts, etc. are detected in these profiles. To distinguish those genes indicative of a response of the cell to perturbation in ergosterol biosynthesis, genes responding in at least five of the eleven experimental treatments were considered in the subsequent analysis. The intention of selecting by these criteria is to identify genes that have a convergent pattern of expression across many individual treatments which may therefore be indicative of a common response.

156 genes showed significant increase in transcript level in five or more treatments and 78 showed significantly decreased transcript level in five or more treatments. These were annotated using the biological role assigned by the Proteome YPD database (Hodges *et al., supra*) and grouped accordingly. The number and characteristics of the responsive genes in these biological role groupings are shown in Table 1. The column headed "number of genes" indicates the number of individual genes in the category responsive in five or more treatments; the column "number of hits" indicates the number of times a gene was responsive to a treatment. The categories in Table 1 are ordered by number of genes. The category with the largest number of hits is the unannotated or "unknown" group of genes, which contains 53 genes. Next most abundant is the group of 36 responsive mitochondrial genes, followed by 22 responsive genes involved in biosynthesis of lipids, fatty acids and sterols. This group includes 9 genes in the ergosterol pathway. The other categories in Table 1 include 24 genes involved in translation and protein processing, 20 stress-response genes, 16 membrane proteins, and a number of other genes involved in other biological pathways. The category "other related genes" refers to genes related through other experimental results to ergosterol perturbation, and are described in Tables 2 and 3 and Figures 5 and 6. The category "other genes" includes responsive genes which were the sole representative of a particular biological pathway, and whose relationship to ergosterol perturbation could not be discerned.

Table 1: Genes Responding in Five or More Treatments

Biological role	# of genes	# of hits	hits/ gene	# of drug hits	Drug hits/ Gene	# of KO hits	KO hits/ gene	Sat	%S	%D	%KO	up/ down
Unknown												
Up	26	154	5.92	124	4.77	30	1.15	0	0	80	20	up
Up/sat'd	8	52	6.50	35	4.38	17	2.13	8	100	67	33	up
Down	11	67	6.09	56	5.09	11	1.00	0	0	84	16	down
Down/sat'd	8	50	6.25	41	5.13	9	1.13	8	100	82	18	down
Mitochondrial	32	194	6.06	145	4.53	49	1.53	23	72	75	25	up
	4	25	6.25	20	5.00	5	1.25	1	25	80	20	down
Lipid/fatty	16	101	6.31	63	3.94	38	2.38	1	6	62	38	up
acid/sterol	6	37	6.16	28	4.67	9	1.50	2	33	76	24	down
Protein												
Synthesis and	15	90	6.00	73	4.87	17	1.13	0	0	81	19	up
Processing	9	54	6.00	42	4.67	12	1.33	6	67	78	22	down
Stress	15	90	6.00	59	3.93	31	2.07	5	33	66	34	up
	5	31	6.20	25	5.00	6	1.20	4	80	81	19	down
Membrane-	8	45	5.62	39	4.88	6	0.75	0	0	87	13	up
Associated	8	50	6.25	36	4.50	14	1.75	3	38	72	28	down
Amino acid biosynthesis	9	58	6.44	43	4.78	15	1.67	7	78	74	26	up
Cell cycle control	7	38	5.43	31	4.43	7	1.00	0	0	82	18	up
Chromatin-associated	6	35	5.83	29	4.83	6	1.00	0	0	83	17	up
Carbohydrate metabolism	6	35	5.83	24	4.00	11	1.83	5	83	69	31	up
Vesicle-associated	5	29	5.80	23	4.00	6	1.20	0	0	79	21	down
Cell wall biosynthesis	5	33	6.60	26	5.20	7	1.40	3	60	79	21	up
Nucleotide metabolism	5	28	5.60	19	3.80	9	1.80	3	60	68	32	down
Other	4	24	6.00	19	4.75	5	1.25	1	25	79	21	down
related genes	3	17	5.67	11	3.67	6	2.00	1	33	65	35	up
Other genes	12	66	5.50	58	4.83	8	0.67	2	17	88	12	down

Table 1 lists genes responding in five or more treatments. Shown are (1) the number of responsive genes in each functional class (biological role), (2) the number of treatments in which these genes responded, (3) the corresponding ratio of hits per gene, (4) the number of drug treatments in which these genes responded, (5) the corresponding average of drug hits per gene, (6) the number of the genetic disruptions with alterations in these genes (# of knockouts (KOs)), (7) the corresponding number per gene, (8) the number of genes which also respond to saturated culture, (9) the percentage of genes responding to saturated culture, (10) the percentage of genes responding to drug treatments, (11) the percentage of genes responding to genetic perturbation, and (12) the direction of the response (increased (up)or decreased (down) expression). Functional annotations were obtained from the YPD database (Hodges *et al.*, *supra*).

Responsive genes are also characterized by the relative number of responses due to chemical vs. genetic perturbation (drug/gene and KO (knockout)/gene, respectively). This is indicated in the columns headed %D and %KO in Table 1. Values ranged from highs of 88-87% of the response due to chemical perturbation, in the case of membrane-associated proteins, to lows of 62% of the response due to response to chemical perturbation, and therefore 38% due to response to genetic perturbation, seen for the class of lipid / fatty acid / sterol pathway genes.

The column headed "Sat", in Table 1, indicates the number of responsive genes in each category whose transcript levels changes as the cells enter saturated growth and undergo the diauxic shift (DeRisi *et al.*, *supra*; Hodges *et al.*, *supra*; G. Bammert and J. Fostel, unpublished). The proportion of responsive genes which are also altered in response to saturated growth varies from 0% of genes in several pathways to a high of 78% and 83% of ergosterol-responsive genes related to amino acid and carbohydrate metabolism, respectively. This behavior was used to subset the unannotated genes in an attempt to further characterize this group of genes. Genes which are responsive to culture saturation have more hits per gene that non-responsive genes, although the data are insufficient to determine if this difference is significant.

Example 3: Responsive Genes in the Ergosterol Pathway

Figure 5 shows the genes involved in the biosynthesis of ergosterol from acyl-CoA, highlighting some of the responsive genes identified above. Genes listed in bold type in Figures 5 and 6 were responsive in this study; those with italicized names showed a decrease in transcript relative to baseline, non-italicized gene names indicate an increase. The number of responsive treatments is indicated with superscripts, where the first number gives the total of the three disruptant strains in which the particular gene responds, and the second number gives the total of the eight drug treatments in which the gene was responsive.

Most notably, nine genes in the ergosterol pathway respond with increased transcript levels to the conditions used in this study. Also responding is NCP1 (NADP-cytochrome P450 reductase), which is the electron donor for squalene epoxidase, lanosterol 14a demethylase and sterol C22 desaturase (Sutter *et al.*, 1989, Biochem. Biophys. Res. Commun 160:1257-1266) and is coordinately regulated with ergosterol genes. The ergosterol pathway had the highest proportion of responsive genes identified, in agreement with previous studies showing that this pathway is the target of azoles and responsive to modulations of ergosterol level.

The biosynthesis of ergosterol, similar to other complex biological processes, involves the coordination of many factors by which the cell regulates the synthesis of an essential component. ERG19 is reported to contribute to regulation of flux through the mevalonate pathway (Berges *et al.*, 1997, J. Bacteriol. 179:4664-4670), and is increased in response to perturbations here. Expression of ERG3 has been shown to increase not only following treatment with antifungal agents (Smith *et al.*, 1996, Mol. Cell Biol. 16:5427-5432) but was also increased by mutations in ERG2, ERG5, and ERG6 (Arthington-Skaggs *et al.*, *supra*). These data are supported with observations reported here with ERG3 showing an increased expression following drug treatment in addition to increased expression in the three ERG deletion strains. Additionally, expression of NCP1 has been shown to increase five fold in a strain constitutively overexpressing ERG11 (Thorsness *et al.*, 1989, Mol. Cell Biol. 9:5702-12), consistent with coordinate increase in transcript levels of these two genes in this study.

One unexpected result relates to ERG9, a branch-point in the isoprenoid pathway. Others have reported that this gene is subject to regulation in strains carrying mutations in ERG3, ERG7, and ERG24, as well as cells treated with ketoconazole (Kennedy *et al.*, *supra*). In contrast to earlier observations, levels of ERG9 transcripts were observed to decrease in four treatments in the current study. This does not meet the criterion of "transcript level altered in five treatments" used to define responsive genes in the current study. Another gene in the pathway, ERG1, was not observed to respond to treatments used here. ERG1 activity appears to be regulated by protein localization (Leber *et al.*, 1998, . Mol. Biol. Cell 9, 375-386) however it has also been reported to respond to oxygen and sterol-limitation (M'baya *et al.*, 1989, Lipids 24 1020-1023). In a study using promoter fusions as a read-out of transcriptional changes, Dimster-Denk *et al.* (1999, J. Lipid Res. 40:850-860) observed the greatest response to fluconazole treatment from ERG2, ERG8, ERG9, ERG12 and ERG19. With the exception of ERG2, these are found in the isoprenoid pathway rather than the ergosterol pathway, and were not observed to respond to treatment used in the current study.

Interestingly, the ergosterol genes were highly responsive to genetic disruption of ERG2, ERG5, and ERG6. Seven of the ten responsive ergosterol genes were increased in all three mutant strains (Figure 5). Transcripts from the three remaining genes, ERG2, ERG5, and ERG6 were increased in two of three mutant strains. Not surprisingly, each was significantly decreased in the strain bearing a knockout of that gene. Responsive genes in other pathways were also altered in all three deletion strains, but this was more sporadic than the uniform response of genes in the ergosterol biosynthetic pathway.

Example 4: Other Responsive Genes Related to Lipid, Fatty Acid and Sterol Biosynthesis

In addition to participating in cell membranes, esterified ergosterol is found in lipid particles, which may serve as storage reservoirs or as intermediates in intracellular transport (Yang *et al.*, 1996, Science. 272:1353-1356; Zinser *et al.*, *supra*). ARE1 is one of the two genes responsible for the esterification of ergosterol, which is the final step in the pathway leading to ergosteryl esters which are accumulated for storage and

maintenance of sterol homeostasis (Yang *et al.*, *supra*). CYB5 (cytochrome B5) was identified by an ability to overcome ketoconazole hypersensitivity when overexpressed in an erg11 background (Truan *et al.*, *supra*). Presently little else is reported about CYB5.

ACH1 (acetyl CoA hydrolase) is specific for acyl CoA molecules formed during the de novo synthesis of saturated fatty acids and may be responding to an overproduction of sterol intermediates formed during inhibition of the pathway. FAS1 (fatty-acyl-CoA synthase) deletion strains have been shown to posses reduced levels of ergosterol esters and sphingolipids indicating a possible role in lipid biosynthesis and metabolism (Daum *et al.*, *supra*). LCB1 (serine C-palmitoyltransferase) is the first enzyme involved with the biosynthesis of the long chain base component of sphingolipids. Increase in transcripts of this transferase suggests a potential interaction between the ergosterol and sphingolipid biosynthetic pathways in yeast.

Transcript levels of several other genes involving lipid, fatty acid, and sterol metabolism were decreased. The SUR2 product hydroxylates the sphingoid C-4 of ceramide (Haak *et al.*, 1997, J. Biol. Chem. 272, 29704-29710), and the decrease observed in this transcript may demonstrate another correlation between sphingolipid and sterol pathways. ELO1 encodes an enzyme responsible for elongation of fatty acids, and decline of this transcript level under conditions of this study may indicate a compensatory response in cellular fatty acid content to limited sterols following perturbation of the ergosterol pathway. OLE1, encoding delta-9 desaturase, is needed for formation of unsaurated fatty acids and also shows decreased transcript level here. OLE1 is repressed by the presence of saturated fatty acids, and thus the decline in OLE1 transcript may indicate an increase in saturated fatty acids, possibly another compensatory response to altered ergosterol. Interestingly this fatty-acid responsive repression of OLE1 is mediated through the FAA1 and FAA4 products (Cobon & Haslam, *supra*), and the level of FAA4 transcripts decreased under the conditions used in this study. The connection of fatty acids to perturbation of ergosterol has not be demonstrated by other studies and may indicate a re-structuring of cell membrane in response to reduced ergosterol.

Example 5: Genes with Products Related to Cell Envelope Structure and Function

While ergosterol is found throughout the cell membranes, it is most abundant in the plasma membrane and secretory vesicles and is reported to play an important role in mitochondrial respiration (Daum *et al.*, *supra*; Parks, *supra*; Zinser *et al.*, *supra*). Depletion of ergosterol with concomitant accumulation of sterol intermediates can result in the alterations of membrane functions, synthesis and activity of membrane bound enzymes, mitochondrial activities, and uncoordinated behavior of the yeast cell (Parks, *supra*; Vanden Bossche, 1985, Curr. Top. Med. Mycol. 1:313-351). Genes encoding cell envelope proteins or products required to modify envelope proteins are included in the responsive subset, as shown in Table 2. Changes in transcript level of these genes may thus represent a response to ergosterol perturbation by treatments used here.

Table 2: Genes Encoding Products Associated with Cell Envelope Structure or Function

Genes involved in protein glycosylation:

ALG5[2,4], ALG3, ALG9, ***ALG6***[3,3], ***ALG8***[1,4], ALG10

transferase (PMT family; **PMT2**[3,2])

CWH41, **GLS2**[1,5], MNS1

Genes encoding O-glycosylated proteins:

AGA1, AGA2, BAR1, CTR1, ***CTS1***[2,3], CWP1, CWP2, DAN1 FET3, FLO1, FLO5, FLO9, FLO10, FUS1, GAS1, GIT1, HKR1, HSP150, KEX2, KNH1, **KRE1**[3,4], ***KRE9***[2,3], MID2, MSB2, PEX15, PGM1, PGM2, PIR1, PIR3, **PRB1**[1,4], SAG1, SEC20, SED1, SED4, SLG1, SRO4, SSR1, STA1, STA2, ***TIP1***[2,5], **TIR1**[3,3], TIR2, YLR110C

Responsive membrane proteins (only responsive genes are shown)

BAP2[2,3], **BAP3**[0,6], ***DIP5***[3,4], ***FET4***[1,7], ***HNM1***[3,5], **HXT1**[0,5], **HXT3**[2,4], **PHO87**[0,5], **RCS1**[1,4], **SMF1**[3,4], ***YDR373W***[1,4], **GR138C**[0,5], **KL146W**[1,4], ***YNL065W***[0,6], **YNL321W**[2,3], ***YOR161C***[2,3], **YOR271C**[2,4]

GPI anchored proteins:

AGa1, AGA1, CWP1, CWP2, **EGT2**[1,4], FLO1, FLO5, FLO9, GAS1, ICWP, **KRE1**[3,4], PRY3, SED1, ***TIP1***[2,5], **TIR1**[3,3], TIR2, YAP3, YCR089W, **YDR055W**[0,5], YDR134C, YDR534C, YEL040W, **YER150W**[0,5], YGR189C, YJR151C, YLR110C, **YNL300W**[3,7], YOR009W, YOR214C

Genes involved in synthesis of β-1,6-glucan:

CWH53, FKS2, KNR4, HKR1, **KRE1**[3,4], KRE5, KRE6, ***KRE9***[2,3], KRE11, SKN1

Genes related to osmotic stress:

Part of osmotic stress response signal pathway: **HOG1**[2,3]

Transcription factor with a role in salt tolerance: ***CIN5***[1,6]

Induced by osmotic stress: **DDR48**[3,4], DDR2, GRE1, **GRE2**[3,4], GRE3, HOT1, PTP2, PTP3 SIP18, SKN7, SKK2, SSK1, STE11, SLN1, SHO1, YPD1

Responsive proteins involved in secretion (only responsive genes are shown):

VID24[0,7], **RET2**[3,4], ***SEC17***[0,5], **COP1**[2,5], ***LST8***[3,4], **NHX1**[0,5], ***YGL054C***[0,6], **YHR138C**[0,5]

Table 2 lists genes encoding products associated with cell envelope structure or function. Genes given in bold type were responsive in the study; italicized bold type indicates genes with decreased transcript level. Superscripts indicate the number of treatments to which the gene responded. The first number in the superscript indicates the number of the three genetic perturbations that elicited a response, and the second number indicates the number of the eight drug treatments that elicited a response. Lists of genes were obtained from YPD database (Hodges *et al.*, 1999) and (Daum *et al.*, *supra*; Haak *et al.*, *supra*; Hamada *et al.*, 1998, Mol. Gen. Genet. 258:53-59; Hamada *et al.*, 1999, J. Bacteriol. 181:3886-3889; Inoue *et al.*, 1995, Eur. J. Biochem. 23:845-854).

Transcript levels of several genes encoding proteins with GPI anchors on the mature product (Hamada *et al.*, 1998, *supra*) are also increased in this study (Table 2). This observation may suggest a compensatory response to alterations in plasma membrane composition following experimental treatments. Alternatively this response may indicate cross-talk between paths of inositol utilization, since inositol is incorporated into both GPI anchors and sphingolipids, and alterations in LCB1 seen here may indicate a regulation of sphingolipid biosynthesis under the study conditions.

A decrease in ergosterol may affect genes associated with the formation and maintenance of cell wall. Two of these, GSC2, encoding a component of β-1,3-glucan synthase (Inoue *et al.*, *supra*), and KRE1, encoding a protein required for β-1,6-glucan assembly (Boone *et al*, 1990, J. Cell Biol.110:1833-1843) may be responding to alterations in cell wall function resulting from a decline in the ergosterol component of the plasma membrane. TIR1, a stress-induced cell wall structural protein, was found to be highly upregulated in all of the deletion mutants and by several drug treatments. TIR1 is a glycoprotein residing in the cell wall that may act to protect the membrane from stress (Kowalski *et al.*, 1995, Mol. Microbiol. 15:341-353). Recently, a link between protein N-glycosylation and the biosynthesis of β-1,6-glucan has been described showing incomplete N-chain processing leads to loss of β-1,6-glucan (Shahinian *et al.*, 1998, Genetics. 149:843-856). Transcripts of several genes encoding products needed for synthesis of N-linked oligosaccharide unit as well as the PMT2 transferase are increased following ergosterol perturbation. These are listed in Table 2, as are a number of responsive genes encoding proteins predicted to be O-glycosylated in the mature form (Hodges *et al.*, *supra*). It is possible that transcript levels of components of the cell envelope and enzymes responsible for modifying them may be altered in response to alterations in the plasma membrane resulting from perturbation of ergosterol biosynthesis.

Example 6: Responsive Mitochondrial Genes

Unlike higher eukaryotes, *S. cerevisiae* has higher levels of ergosterol in the mitochondrial inner membrane than other mitochondrial membranes (Tuller & Daum, 1995, FEBS Letters 372:29-32; Zinser *et al.*, *supra*). Previous work has shown that changes in sterol composition of membranes can affect membrane function (Cobon & Haslam, *supra*; Parks, *supra*), and it could be predicted that perturbations of ergosterol levels within the cell may affect the function of mitochondrial enzymes. This pattern does indeed emerge with the increase in transcript level of several components of the mitochondrial electron transport system, as shown in Table 3.

Table 3: Genes Encoding Mitochondrial Proteins

Electron transport complexes, inner mitochondrial membrane:

Ubiquinol cytochrome c reductase complex III

COB, **CYT1**[2,4], **COR1**[2,3], QCR2, **QCR6**[1,4], QCR7, 8, 9, 10, **RIP1**[1,5]

Cytochrome c (anaerobic isoform)

CYC7[1,5]

Cytochrome c oxidase

COX1, COX2, COX3, **COX4**[3,4], **COX5A**[2,4]/ ***COX5B***[1,8],

COX6, COX7, **COX8**[0,5], COX9, COX12, **COX13**[3,4]

ATP synthase

ATP1[3,5], ATP2, ATP3, ATP4, ATP5, ATP6, ATP7, ATP8, ATP9,

ATP14[0,6], **ATP15**[3,4], **ATP16**[0,5], ATP17, ATP20, **INH1**[0,6], TIM11

ADP / ATP carrier protein

AAC1, ***AAC3***[1,4], **PET9**[3,3], **YPL134C**[1,4], YPR021C

Elements of the TCA cycle:

CIT1, **ACO1**[2,4], IDH1, IDH2, **KGD1**[1,5], KGD2, LPD1, LSC1, LSC2,

SDH1[2,4], SDH2, SDH3, SDH4, **FUM1**[2,4], MDH1

Other responding mitochondrial proteins:

ARG5,6[0,5], **ARG7**[1,6], **COT1**[3,5], ***GPD2***[3,4], **IDP1**[0,5], ***ILV5***[2,3], **ILV6**[3,5], **MRP1**[0,6],

PFK27[0,6], **SHM1**[1,4], **SOM1**[0,6], **SOD2**[2,4], **TIM23**[3,3], **YAH1**[2,3], **YIL154C**[0,6]

Table 3 lists genes encoding mitochondrial proteins. Genes given in bold type were responsive in the study; italicized bold type indicates genes with decreased transcript level. Superscripts indicate the number of treatments to which the gene responded. The first number in the superscript indicates the number of the three genetic perturbations that elicited a response, and the second number indicates the number of the eight drug treatments that elicited a response. Lists of genes were obtained from the ExPASy / Boehringer Mannheim metabolic pathways (http://www.expasy.hcuge.ch/cgi-bin/search-biochem-index) and from YPD (Hodges *et al.*, *supra*).

Transcript levels of four members of the cytochrome c oxidase complex, COX4, COX5A, COX8, and COX13 were also increased. Cytochrome oxidase resides in the mitochondrial inner membrane and catalyzes the transfer of electrons from reduced

cytochrome c to molecular oxygen. Similarly, transcript levels were increased for four members of cytochrome c reductase, RIP1, CYT1, QCR6, and COR1, a component of mitochondrial respiration which catalyzes the transfer of electrons to cytochrome c with concomitant transfer of protons across the inner mitochondrial membrane. Another critical respiratory enzyme residing in the mitochondrial inner membrane is ATP synthase. Five genes encoding subunits of the enzyme were observed to be upregulated (ATP1, ATP14, ATP15, ATP16, and INH1). Other genes involved with energy generation whose transcripts were increased are listed in Table 3, as are a number of other responsive genes encoding mitochondrial proteins, including YAH1, which contains a cytochrome c heme-binding signature. (YPD; Hodges *et al.*, *supra*).

Perturbation of mitochondrial electron transport could arise from a decrease in ergosterol in the inner membrane due to interference of the treatments with ergosterol biosynthesis, or from a direct interaction between the chemical agents used and the mitochondrial enzyme complexes. For example, the cytochrome c complex has been analyzed with the use of metal-chelating drugs, which block reduction at the ubiquitinol oxidation site (Boumans *et al.*, 1997, J. Biol. Chem. 272:16753-16760). Since more than half of the responsive mitochondrial genes are found to respond in both genetic alterations and in response to drug treatment, it is likely that the effect is mediated through ergosterol biosynthesis and does not arise as a direct consequence of drug action.

Example 7: Response to Oxygen Stress

Transcripts from four of the five members of the hypoxic gene family (ANB1, COX5b, CYC7, and HEM13) are reduced in response to treatments used in this study (Table 4). Two of these genes, CYC7 and COX5b, are responsible for encoding the hypoxic isoforms of cytochrome c and cytochrome c oxidase. Their expression level has been shown to depend on the level of available oxygen (Burke *et al.*, 1997, J. Biol. Chem. 272:14705-14712). In addition, expression of HEM13 is repressed by oxygen and heme (Keng, 1992, Mol. Cell Biol. 12:2616-2623). Thus the decrease in transcript level seen for these anaerobic-induced genes may be in response to increased levels of intracellular oxygen. Another piece of evidence to support this hypothesis is the observation that

transcripts from four genes involved with oxidative stress response (AHP1, GRE2 YDR453C, and SOD2) are also increased (Table 4).

Table 4: Genes Responding to Oxygen Stress

Genes induced by hypoxia:	AAC1, ANB1, ***COX5B***[1,8], CPR1, ***CYC7***[1,5], **ERG11**[3,3], ***HEM13***[0,7], HMG2, ***OLE1***[2,7], ROX1, SUT1
Genes induced by anarobiasis:	CYB2; **ERG11**[3,3], ***TIP1***[2,5], TIR2, **TIR1**[3,3],
Major oxidant scavenging enzymes:	CCP1, CTT1, TSA1, SOD1, **SOD2**[2,4], TRR1, TRX1, RX2, GLR1, **YDR453C**[2,4], YCL035, AHP1, **GRE2**[3,4]
Pentose phosphate pathway shunt:	**ZWF1**[3,4]

Table 4 lists genes encoding proteins responsive to oxygen stress. Genes given in bold type were responsive in the study; italicized bold type indicates genes with decreased transcript level. Superscripts indicate the number of treatments to which the gene responded. The first number in the superscript indicates the number of the three genetic perturbations that elicited a response, and the second number indicates the number of the eight drug treatments that elicited a response. Lists of genes were obtained from YPD (Hodges *et al.*, *supra*) and from (Jamieson, 1998, Yeast 14, 1511-15; Slekar *et al.*, 1996, J. Biol. Chem. 271:28831-28836; Zitomer *et al.*, 1997, Kidney International 51:507-513).

Recently, Kwast *et al.* (1999, Proc. Natl. Acad. Sci. USA. 96:5446-5451), have determined that OLE1, CYC7, and COX5b belong to a separate class of hypoxia-responsive genes whose regulation is sensitive to carbon monoxide. These authors have concluded that cytochrome c oxidase may act as an oxygen sensor for the induction of hypoxic genes. In the course of normal catalysis, cytochrome c oxidase transfers two electrons to molecular oxygen; incomplete reactions can result in the formation of reactive

oxygen species. Machida *et al.* (1998, J. Bact. 180:4460-4465) have shown that exposure to farnesol results in the generation of reactive oxygen species through an indirect effect on mitochondrial electron transport. Farnesol can be generated from farnesyl pyrophosphate, a precursor to sterols. It is possible that limitation of ergosterol could also impact the activity of cytochrome c oxidase or other components of the electron transport chain, leading to the generation of reactive oxygen species.

Example 8: Heme Regulation

Transcripts of HEM1, the first gene involved in the biosynthesis of heme, increased in response to eight of the eleven treatments tested. Interestingly, all eight were drug treatments, suggesting a compensatory response to drug treatment not represented in the mutant strains. It is possible that this results from heme depletion by the drugs, or from an unanticipated secondary effect of drug exposure. Transcript levels of HEM13, the rate-limiting step of heme biosynthesis (Zitomer & Lowry, 1992, Microbiol. Rev. 56, 1-11) were reduced in response to seven drug treatments and again, in none of the mutational perturbations. HEM13 has been shown to be repressed by heme and oxygen (Amillet *et al.*, 1996, J. Biol. Chem. 271:24425-24432; Keng, *supra*), suggesting that these may be limiting under the conditions of the drug treatments.

Heme plays a central role in sterol synthesis and has been shown to transcriptionally regulate several genes involved with the process (Parks & Casey, 1995, Ann. Rev. Microbiol. 49:95-116; Thorsness *et al.*, *supra*). The accumulation of 5-aminolevulenic acid, the product of Hem1p, derepresses 3-hydroxy-3-methyl-glutaryl CoA reductase, leading to increased levels of 2,3-oxidosqualene (Lorenz *et al.*, *supra*). Heme is also required for the enzymatic activity of Erg3p (C-5 sterol desaturase), Erg5p (C-22,23 desaturase) (Parks, *supra*). Erg11p contains heme, and shows heme-regulated expression (Turi & Loper, *supra*).

Heme also plays a role in sensing intracellular oxygen levels (Zitomer & Lowry, *supra*). Genes regulated by heme are listed in Figure 6. In the present study, transcripts from four heme-induced genes are increased and transcripts from three heme-repressed genes are decreased, consistent with induction of heme regulation of expression.

Interestingly, ERG11, another heme-repressed gene, is increased, however this is likely due to additional regulation of this gene by ergosterol levels, as discussed above.

Because transcripts of HEM1 and HEM13 in the heme pathway respond only in cells treated with chemical agents, it is possible that levels of heme are directly affected by chemical interaction with the agents used. This prediction would be consistent with alterations in expression of heme-containing proteins. Heme containing proteins identified by the YPD data base (Hodges *et al.*, *supra*) are listed in Figure 6. Transcripts of eight genes encoding heme-containing proteins are responsive to the conditions used, although two are decreased and six are increased. Inspection of the responsive heme-containing proteins reveals that they are members of groups regulated by other factors, including oxygen level and heme level. Thus their behavior does not shed additional light on effects of azoles and other agents used here on other heme-containing molecules in the cell.

Example 9: Correlation of the Expression Pattern of a Novel Azole to Other Conditions

Included in this analysis was the expression profile in response to treatment with a compound containing an azole moiety identified in a whole cell screen of *S. cerevisiae*. PNU-144248 contains an imidazole ring, yet is structurally distinct from the other azoles tested. The rationale for including it in this study is to determine if exposure to this unknown compound results in a pattern of expression similar to that seen in response to treatment with azoles of known biochemical function, thereby suggesting a similar mode of action. Such information is useful to assess the predictive ability of expression profiles observed in response to an agent with an unknown mode of action. Of the 156 treatment-affected genes with increased transcript levels following ergosterol perturbation, 144, or 92%, were increased also in response to treatment with PNU-144248. All of the ergosterol, lipid, fatty-acid and sterol metabolism genes and 17 of the 19 genes involved with energy generation were included (the exceptions were QCR6 and COX13). 29 of the 34 of the unknowns were also included. From the set of 78 treatment-affected genes with decreasing transcript levels following treatment, 40, or 51%, also decreased following treatment with PNU-144248. These data suggest PNU-144248 does indeed behave as an

azole as measured by cellular responses at the level of gene transcription.

It is interesting that a greater proportion of genes with increased transcript level were in common between PNU-144248E and other treatments known to perturb ergosterol biosynthesis than genes with decreased transcript level. While the data are not sufficient to determine the significance of this difference, it is consistent with the overall patterns seen in Figures 5 and 6 and Tables 2-4, which contain a preponderance of genes with transcript level increased in response to the treatments used here and relatively few genes with decreased transcript level. A number of factors may contribute to this. Genes described in the tables and figures were selected because they have been characterized in other studies, and it is possible that positive changes in transcript level were more likely to be detected in these studies than a decrease in level might have been. There are also more responsive genes included in the class with increased transcripts, so it is perhaps not surprising that numerically more were included in the figures. Finally, it is more straightforward to measure an increase in a biological product if the level is changing due to new synthesis, as would be expected if the level were increasing. The turnover of particular transcripts is not well characterized, and differences in transcript levels observed may reflect stability and turnover rate in addition to nascent synthesis, especially in transcripts seen to decline following treatment.

Example 10: Genes of Unknown Function

Additional genes responsive to perturbation of the ergosterol pathway have been identified by the methods of this invention. Of the 156 genes found to be upregulated by the criteria used herein, 34 are of unknown function and little is known about the majority of these genes. Interestingly one of these, COS8, is increased in ten of the eleven conditions tested and another, AFG1, is increased in nine conditions. The only other gene responding in more than eight conditions was YNL300W, encoding a cell wall component with putative GPI anchor (Hamada *et al.*, 1999, *supra*). These genes, in conjunction with several other highly responsive unknown genes, represent a novel class of genes that may play a role in the cellular response to a decline in ergosterol or perturbation of biosynthetic machinery. These highly responsive unknowns may be indicative of a key compensatory

response that, if targeted, could be a point of chemical intervention that would be deleterious to the fungal cell.

The following genes, for which some biological information is available, have increased transcript levels:

CLN3, CPA1, DAK1, DOT6, DSK2, ENO2, HOM3, HSP104, LYS20, MET17, MET6, NTH1, NUD1, PAT1, PDI1, PRY1, PYC2, RPC19, RPO21, RPS18B, RPS26A, RRP42, SAH1, SAM2, SCC3, SCH9, SPE1, SRA1, SSA1, STP4, TUP1, YLR109W, YMR173W, YMR304C, YNL065W, YPR184W

The following genes of unknown function show increased transcript levels:

AFG1, COS8, HMO1, PTC2, YAR068W, YBL081W, YBR269C, YCL042W, YDL012C, YDL039C, YDL204W, YDR157W, YDR247W, YDR433W, YDR476C, YEL007W, YER035W, YFR017C, YGR165W, YGR294W, YHB1, YHR039C, YLR100W, YLR251W, YLR356W, YML030W, YMR310C, YNL134C, YNR046W, YOL027C, YOR248W, YOR267C, YOR359W, YPL014W, YPL017C, YPL067C, YPL247C, YPR144C, YPR157W

The following genes, for which some biological information is available, show a decrease in transcript level:

ADE1, APT1, CMK2, DED1, ECM1, GAR1, HIS5, HSP26, HSP42, HTA2, MRT4, OYE2, PHO11, PHO5, PKH1, RAS2, RKI1, RNR4, RPL39, RRPS25, SML1, STU1, TIF51B, URA4, YAR073W, YBL005W-b, YBR012W-a, YBR204C, YDR516C, YLL025W

The following genes of unknown function respond with decreased transcript levels: YAL008W, YAR003W, YBR113W, YCL036W, YDR214W, YDR492W, YEL033W, YER067W, YER138C, YER160C, YGL020C, YGR243W, YHL021C, YJL012C, YJL055W, YJR027W, YLR065C, YML039W, YML040W, YMR051C, YMR316W, YNL087W, YOL002C, YOL014W, YOR175C, YPL019C

Example 11: Changes in Transcript Level Over Time

It is likely that responses occurring either earlier or later than the time of exposure chosen for microarray analysis were not detected. Quantitative polymerase chain reaction

(PCR) reactions were used to measure transcript levels from a few genes in the ergosterol pathway. This was done to define the time frame of the cellular response to perturbation of ergosterol by chemical agents and to confirm changes in transcript levels from responsive genes identified by microarray hybridization.

Cultures were grown in the presence of two azoles, ketoconazole and voriconazole, and aliquots harvested at 1, 3, 5, and 24 hrs of exposure. Total RNA was extracted as described above and used as a substrate for quantitative PCR, using probe/primer sets to detect ERG3, ERG4, ERG5, ERG6, ERG11 and ERG24 transcripts. PCR reactions were performed using the Taqman Gold PCR kit and analyzed using a Prism 7700 (PE Applied Biosystems).

In this assay, the number of reaction cycles required to raise the level of product above a threshold (C_t) is a measure of the abundance of the initial template. Threshold values are calculated from 10 standard deviations of background. To control for minor variance in loading conditions, each PCR reaction also included an internal control reaction, to translation elongation factor TEF1, which has relatively constant expression levels that are not predicted to vary in response to perturbation of ergosterol (data not shown). Data reported for each reaction are given as the difference between experimental C_t and the C_t of the standard gene, TEF1. The difference between experimental C_t and control C_t is a measure of the change in abundance of experimental transcript in each reaction. These values are plotted in Figure 7. An increase in cycle number indicates a decrease in transcript abundance. Examination of Figure 7 reveals that the response of cells to all four exposures was very similar for the five transcripts measured.

Probe-primer sets for the ERG3, ERG4, ERG5, ERG6, ERG11 and ERG24 were identified using Primer Express (PE Applied Biosystems, Foster City, CA). Sequences of primers and fluorescently labeled probes were:
5'-CCAAATGGCAATGGAAATCAA (SEQ ID NO:1), 5'-
ACAAACCATGGAACGGTCAAC (SEQ ID NO:2), 6FAM-
TTGGCAGTCAGTGCTATCCCATGGATG-TAMRA (SEQ ID NO:3), ERG3;
5'-TCACTTTGGGTGCCTGTTTG (SEQ ID NO:4), 5'-
ACATGACAACCCCCAATTGTG (SEQ ID NO:5), 6FAM-

TGTCACATAGCCGTAAGTCTCCCACTGCT-TAMRA (SEQ ID NO:6), ERG4;5'-CGTGCATCAACTTACACCAAGC (SEQ ID NO:7), 5'-GTGCTCAAATGGCTAAGGATCAT (SEQ ID NO:8), 6FAM-TAACACCAAACTGGCTTACCACCTGCAGCAA-TAMRA (SEQ ID NO:9), ERG5; 5'-AGACACCTACAGAGGAGGAATCACA (SEQ ID NO:10), 5'-GACCACAGATCTCTCAACTACTTCAGA (SEQ ID NO:11), 6AM-TGCAACCCCAACTCCCAAACCTTTAAGT-TAMRA (SEQ ID NO:12), ERG6; 5'-CCACCTCTAGTGTTTTACTGGATTCC (SEQ ID NO:13), 5'-CTTCGAAAAACTCGTATGGCTTC (SEQ ID NO:14), 6FAM-TACCGTACACAACAGCACTACCGACCCA-TAMRA (SEQ ID NO:15), ERG11; 5'-ACCACAGGCAATAGACAGTCCATA (SEQ ID NO:16), 5'-TGAGCTTTGGAACGGTATCAAG (SEQ ID NO:17), 6FAM-TTCACGATTGCCCAGATAGTGTAGCGCA –TAMRA (SEQ ID NO:18), ERG24.

Primers and probes used for TEF1 were: 5'-GTAGAGTTGAAACCGGTGTCATCA (SEQ ID NO:19), 5'-AACGGACTTGACTTCAGTGGTAACA (SEQ ID NO20), VIC-CAGGTATGGTTGTTACTTTTGCCCCAGCTG-TAMRA (SEQ ID NO:21).

Two different probe/primer sets were used to measure ERG11 transcripts. These are labeled ERG11 and ERG16 (reflecting alternative names for the same gene) in Figure 7. The behavior of these two lines in each plot reveals the qualitative nature of this experiment, while the ERG11 and ERG16 response curves were parallel in each experiment they differ in magnitude. This most likely reflects differences in the efficiencies of the two probe/primer sets. Because of this, the shape of each curve was considered more significant than its absolute magnitude.

Transcripts from ERG3 increase throughout the time course. Other transcripts show a transient decrease (as reflected by the upward deflection of the curve, reflecting higher cycle number to reach threshold level relative to the internal control). This is followed by a recovery, and possible increase in level of transcript over the period

between 5 to 24 hours. Based on this, the microarray samples, taken at 90 minutes of exposure, are likely to capture much of the cellular response, at least as reflected by the behavior of the set of five genes measured here.

Example 12: Description of a database relating functional annotation to expression profile

Microarray experiments can reveal the transcriptional profile of a cell under a given condition. To interpret this profile it is helpful to analyze the pattern of responsive genes in the context of the biological pathways altered in response to treatment. Towards this goal a relational database was constructed to connect gene annotations with designations on the array. Gene designations from the array were indexed to standard names for each gene and the biological role designation for each gene, indexed by standard name, was obtained from a licensed database of the yeast literature, YPD. Two tables were created in the database: (1) a table relating a biological role to each gene on the Affymetrix yeast microarray and (2) a table relating a chip feature to each annotated gene. The microarray manufacturer provided the first table and the second was constructed by eliminating unannotated genes and genes designated to be of unknown function, then making multiple entries for genes fulfilling multiple biological roles, and, finally, combining related biological roles into groups relating to a particular pathway. These tables allow one to determine which biological pathways are most responsive to a particular treatment.

The YPD database provides a service that allows relational databases to connect transcript profiles with functional annotations. The YPD database provides this service. Additionally, descriptions of such databases have been published by several labs. Recent examples include:

Systematic management and analysis of yeast gene expression data.
Aach J, Rindone W, Church GM; Genome Research 10, 431-45, 2000
Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115 USA, and the Lipper Center for Computational Functional Genomics, Boston, Massachusetts 02115.

Mining and visualizing large anticancer drug discovery databases.
Shi LM, Fan Y, Lee JK, Waltham M, Andrews DT, Scherf U, Paull KD, Weinstein JN;
J. Chem. Inf. Conput. Sci. 40, 367-79, 2000
Laboratory of Molecular Pharmacology, Division of Basic Sciences, Information
Technology Branch, National Cancer Institute, Bethesda Maryland 20892.

Example 13: Pathway Annotation

S. cerevisiae cells were exposed to a variety of treatments and responsive genes identified using the method described above for perturbations of ergosterol biosynthesis. Responsive genes for each treatment were flagged in the database. The table of flagged genes was related to the annotated genes in the database and a histogram developed of genes in each pathway responsive to each treatment. An example of such a histogram is provided in Figure 8. This method has the advantage that the same subset of genes is queried for each treatment, and the magnitude of response from each pathway can be compared without knowledge of the total number of potentially responsive genes in each pathway. Such a histogram allows similarity among treatments to be discerned and the formation of hypotheses about the pathway or pathways responsive to a particular treatment.

Example 14: Mode of Action Prediction

This method was applied to members of the indazole class of antifungal agents to predict the mode of action of this class of compounds. Cells were exposed to half-MIC concentrations of four analogs of 3-phenylindazole previously identified by Phil Zaworski and Marek Nagiec and synthesized by Fred Ciske and Michael Genin. Cells were grown in rich medium at 30 degrees with shaking at 225 rpm. Cells were harvested following either 1 or 3 hours of exposure, RNA extracted and labeled as described above, and hybridized to Affymetrix yeast microarrays.

In this example data were normalized to transcript profiles of untreated cells harvested at either 1 or 3 hours. The intensity measure for each gene was normalized for chip variation by division with the geometric mean of the distribution of expressed genes

on the particular chip, as described above for the ergosterol example. Genes whose transcript level increased in response to each of the four treatments at either or both time points were considered to be indicators for that treatment class. This set of responsive genes consisted of 350 genes.

The set of 350 genes with responsive transcripts contained 68 transcripts whose product is related to cell envelope integrity and biosynthesis, 65 transcripts related to bud formation, vesicular transport and DNA biosynthesis/checkpoint control, 79 transcripts related to osmotic and chemical stress. Taken together, this profile is consistent with loss of envelope integrity resulting from perturbation of microfilament transport or bud formation. This pattern is consistent with other tests for biological activity of this class of antifungals performed here (data not shown).

In addition to annotated genes, the set of responsive transcripts contained 92 genes which were unannotated at the time of this analysis. These are candidates for additional components of the indazole-response. Further, this set may contain the molecular target or additional biologically-related targets of indazoles. The set of unannotated indazole-responsive genes is given below:

YAR002W	YBR007C	YCL047C	YCL049C	YCR061W
YCR062W	YCR076C	YCR095C	YDL099W	YDL173W
YDL177C	YDL241W	YDR163W	YDR383C	YDR391C
YDR407C	YDR428C	YDR466W	YDR482C	YDR504C
YDR533C	YER030W	YER037W	YER128W	YFR003C
YGL072C	YGL108C	YGL117W	YGL174W	YGR016W
YGR035C	YGR042W	YGR131W	YGR189C	YGR243W
YHR059W	YIL024C	YIL117C	YJL021C	YJL037W
YJR054W	YJR079W	YJR088C	YKL052C	YKL054C
YKL076C	YKR011C	YKR089C	YLR011W	YLR099C
YLR100W	YLR194C	YLR198C	YLR346C	YLR414C
YML041C	YML053C	YML131W	YMR010W	YMR031C
YMR098C	YMR103C	YMR130W	YMR136W	YMR184W
YMR245W	YNL133C	YNL149C	YNL212W	YNL258C

YNR025C	YNR067C	YOL014W	YOL027C	YOL034W
YOL101C	YOL150C	YOR013W	YOR049C	YOR051C
YOR152C	YOR189W	YPL034W	YPL052W	YPL067C
YPL070W	YPL071C	YPL107W	YPL156C	YPL168W
YPR105C	YPR158W			

All references cited herein are incorporated herein in their entirety by reference.

We claim:

1. A method of preparing an expression profile of genes expressed by a cell treated with an agent, comprising the step of plotting the log(ratio) of normalized filtered ADI values resulting from a hybridization assay performed on a microarray with mRNA extracted from the cell against a sum of the normalized filtered ADI values.

2. The method of claim 1, wherein the agent is an antibiotic agent.

3. The method of claim 2, wherein the antibiotic agent is an antifungal agent.

4. The method of claim 1, wherein the agent has an unknown mode of action.

5. A method of identifying an agent capable of altering expression of a gene in a cell, comprising the step of:

identifying a treatment-affected gene from a prepared expression profile, wherein the treatment-affected gene's datapoint on the expression profile exceeds one standard deviation from a distribution.

6. The method of claim 4, wherein the datapoint exceeds two standard deviations from the distribution.

7. A method of identifying a first agent capable of interfering with a biosynthetic pathway, comprising the steps of:

comparing a plurality of first treatment-affected genes identified from an expression profile of a first cell treated with the first agent, with a plurality of second treatment-affected genes identified from an expression profile of a second cell treated with a second agent known to interfere with the biosynthetic pathway; and

determining whether the number of the first treatment-affected genes which are identical to the second treatment-affected genes exceeds a percentage of about 20% of the total number of second treatment-affected genes.

8. The method of claim 7, wherein the percentage exceeds about 30%.

9. The method of claim 7, wherein the percentage exceeds about 40%.

10. The method of claim 7, wherein the percentage exceeds about 50%.

11. The method of claim 7, wherein the percentage exceeds about 60%.

12. The method of claim 7, wherein the percentage exceeds about 70%.

13. The method of claim 7, wherein the percentage exceeds about 80%.

14. The method of claim 7, wherein the percentage exceeds about 90%.

15. The method of claim 7, wherein the percentage is about 100%.

16. A method of identifying a gene as part of a biosynthetic pathway comprising the step of:

identifying a treatment-affected gene common to a percentage of expression profiles prepared from cells treated with agents known to affect the biosynthetic pathway.

17. The method of claim 16, wherein the percentage is at least about 25%.

18. The method of claim 16, wherein the percentage is at least about 40%.

19. The method of claim 16, wherein the percentage is at least about 55%.

20. The method of claim 16, wherein the percentage is at least about 75%.

21. The method of claim 16, wherein the percentage is at least about 85%.

22. The method of claim 16, wherein the percentage is about 100%.

23. The expression profile prepared by the method of claim 1.

24. The expression profile of claim 23, wherein the cell is a yeast cell, and wherein the agent is a compound selected from the group consisting of azole compounds and imidazole compounds.

25. An agent identified by the method of any of claims 4 or 6.

26. A gene, in isolated form, identified by the method of claim 15.

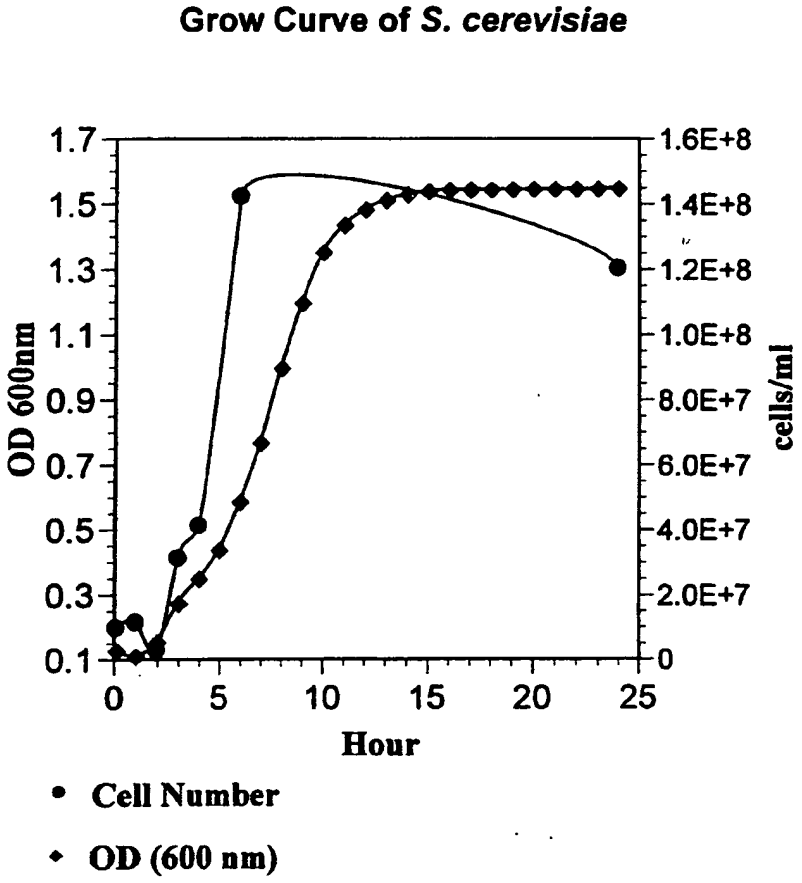

Grow Curve of S. cerevisiae
OD 600nm
cells/ml
Hour
1.7
1.5
1.3
1.1
0.9
0.7
0.5
0.3
0.1
1.6E+8
1.4E+8
1.2E+8
1.0E+8
8.0E+7
6.0E+7
4.0E+7
2.0E+7
0
0
5
10
15
20
25
Cell Number
OD (600 nm)

FIGURE 1

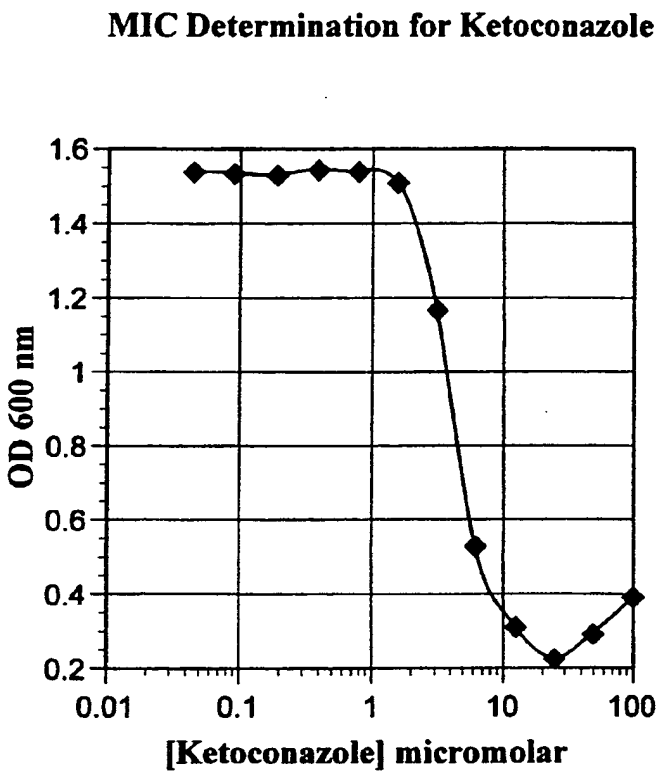

MIC Determination for Ketoconazole
OD 600 nm
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0.01
0.1
1
10
100
[Ketoconazole] micromolar

FIGURE 2

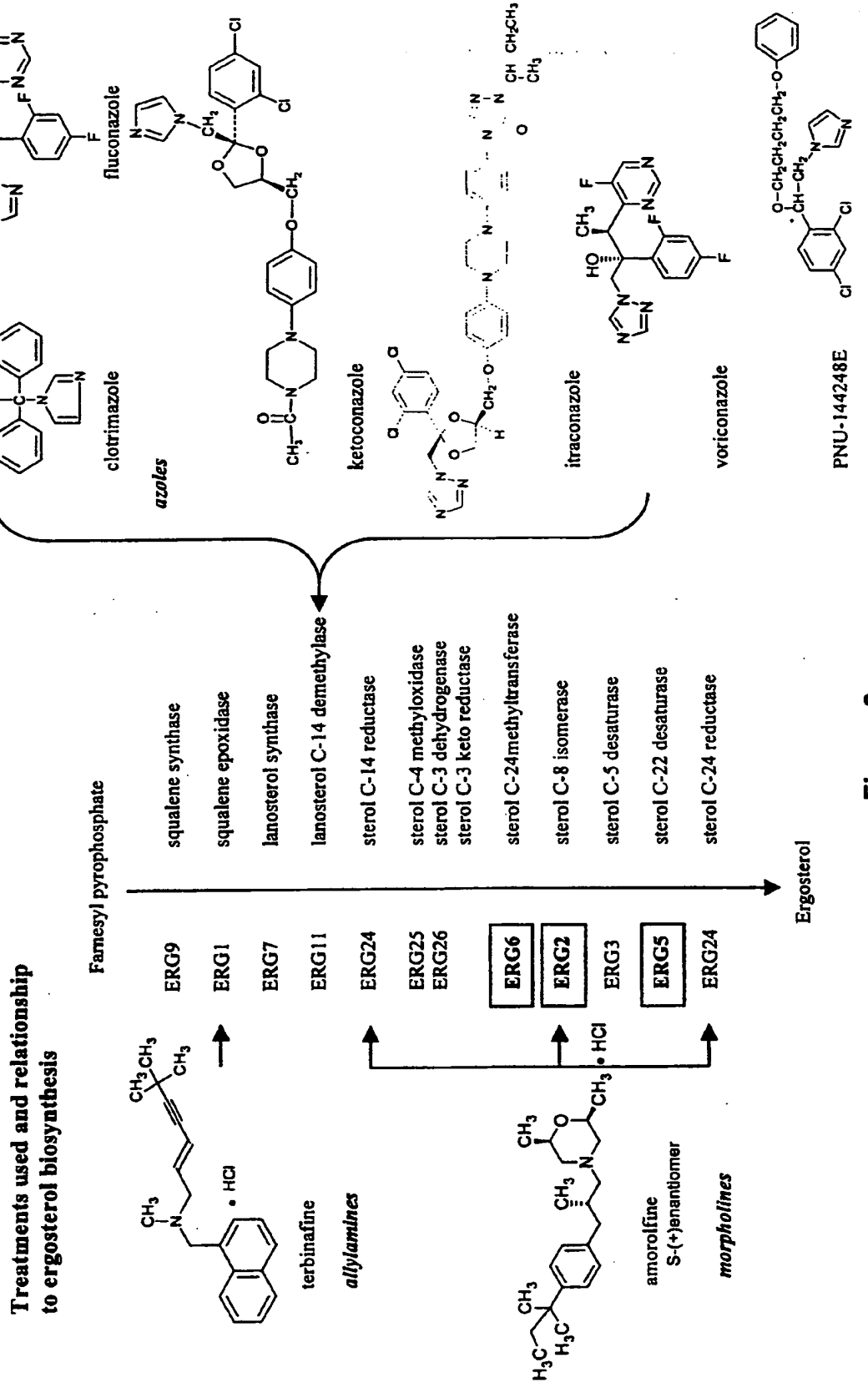

Treatments used and relationship to ergosterol biosynthesis
Farnesyl pyrophosphate
squalene synthase
squalene epoxidase
lanosterol synthase
lanosterol C-14 demethylase
sterol C-14 reductase
sterol C-4 methyloxidase
sterol C-3 dehydrogenase
sterol C-3 keto reductase
sterol C-24methyltransferase
sterol C-8 isomerase
sterol C-5 desaturase
sterol C-22 desaturase
sterol C-24 reductase
Ergosterol
ERG9
ERG1
ERG7
ERG11
ERG24
ERG25
ERG26
ERG6
ERG2
ERG3
ERG5
ERG24
• HCl
terbinafine
allylamines
amorolfine
S-(+)enantiomer
morpholines
azoles
clotrimazole
fluconazole
ketoconazole
itraconazole
voriconazole
PNU-144248E

Figure 3

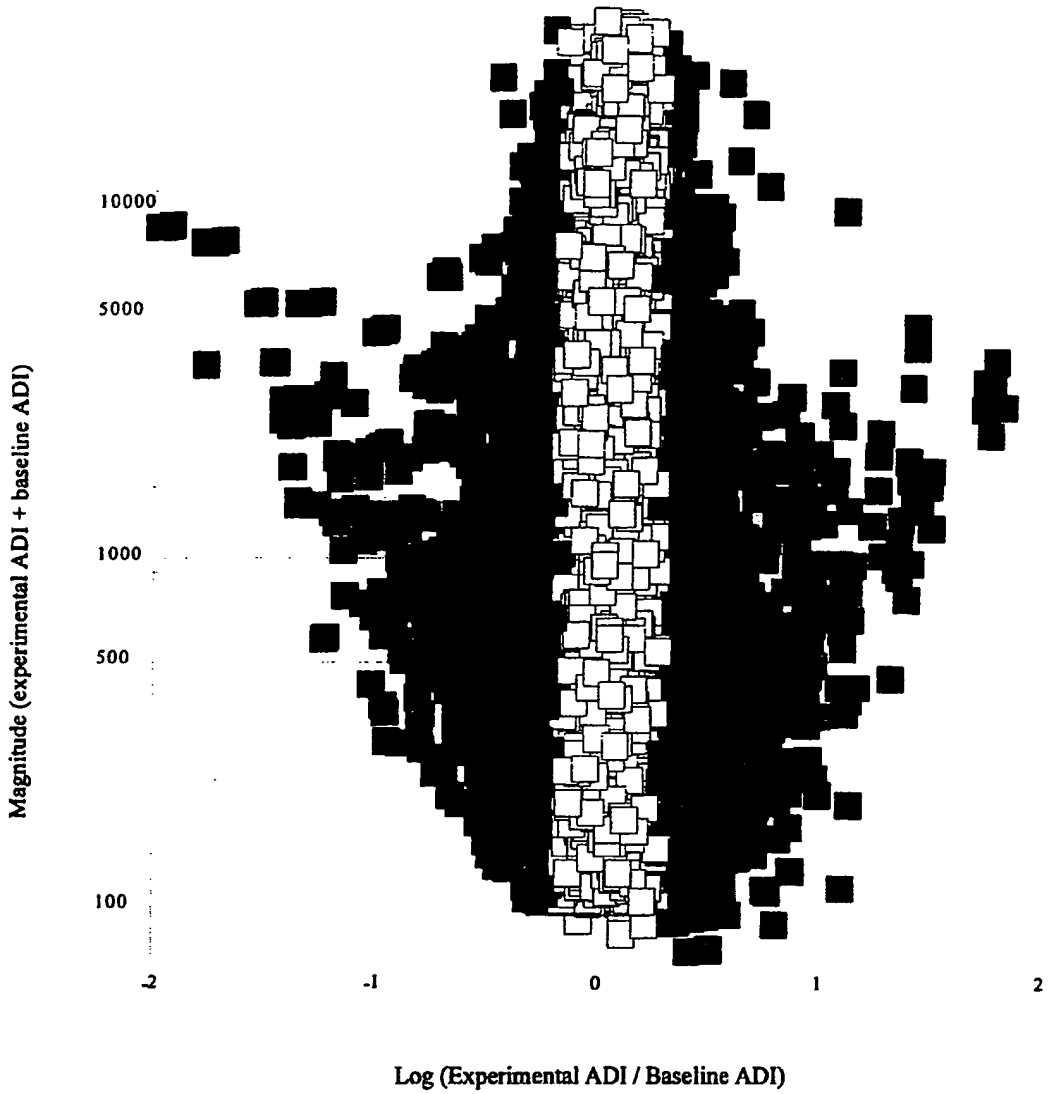

10000
5000
1000
500
100
Magnitude (experimental ADI + baseline ADI)
-2
-1
0
1
2
Log (Experimental ADI / Baseline ADI)

Figure 4

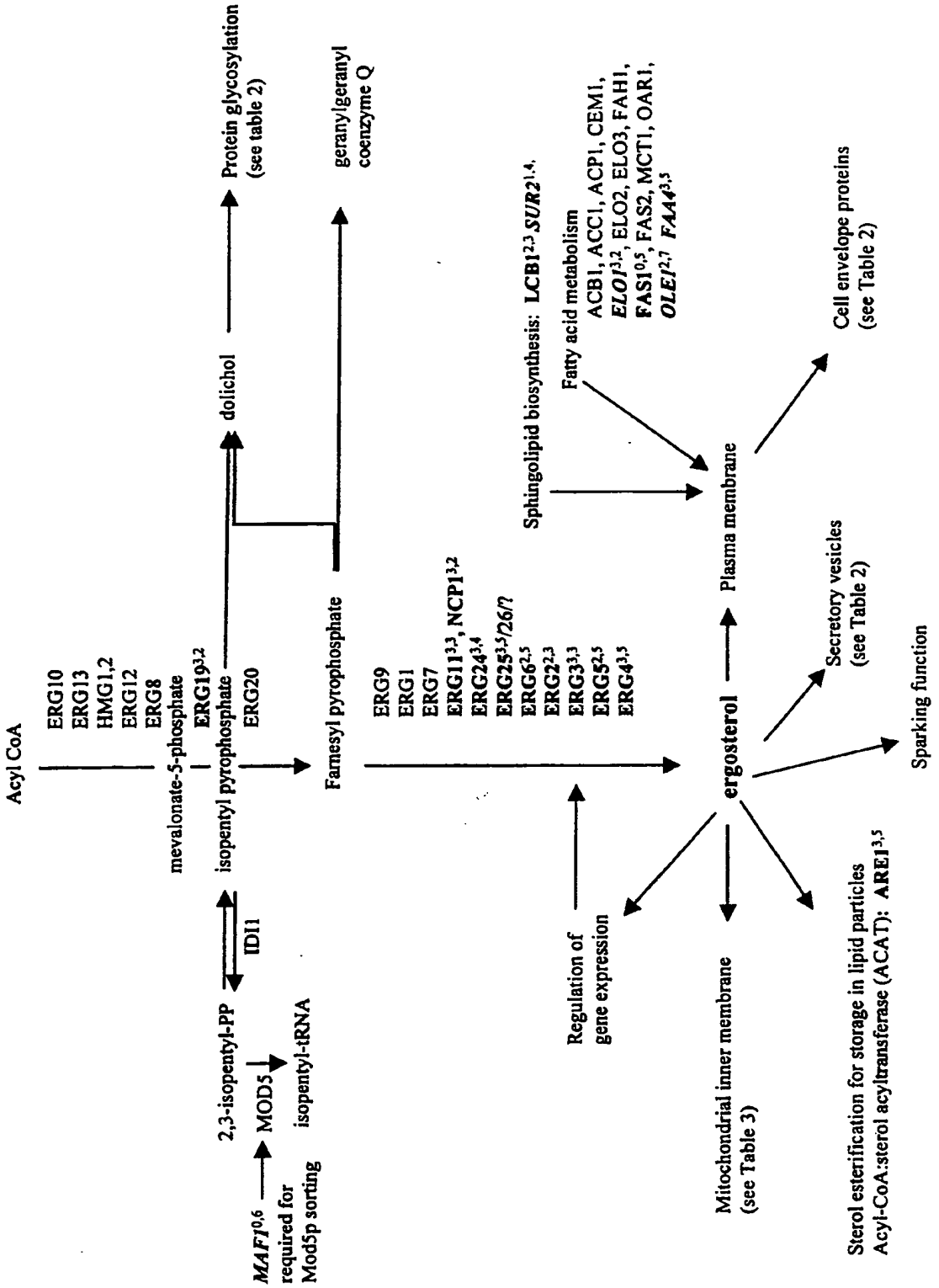

Acyl CoA
ERG10
ERG13
HMG1,2
ERG12
ERG8
mevalonate-5-phosphate
ERG19[3,2]
isopentyl pyrophosphate
ERG20
Farnesyl pyrophosphate
dolichol
Protein glycosylation (see table 2)
geranylgeranyl coenzyme Q
ERG9
ERG1
ERG7
ERG11[1,3], NCP1[3,2]
ERG24[3,4]
ERG25[3-5]/26/7
ERG6[2,5]
ERG2[2,3]
ERG3[3,3]
ERG5[2,5]
ERG4[3,5]
Sphingolipid biosynthesis: LCB1[2,3] SUR2[1,4]
Fatty acid metabolism
ACB1, ACC1, ACP1, CEM1,
ELO1[3,2], ELO2, ELO3, FAH1,
FAS1[0,5], FAS2, MCT1, OAR1,
OLE1[2,7] FAA4[3,5]
Plasma membrane
Cell envelope proteins (see Table 2)
ergosterol
Secretory vesicles (see Table 2)
Sparking function
IDI1
2,3-isopentyl-PP
MOD5
isopentyl-tRNA
MAF1[0,6] required for Mod5p sorting
Regulation of gene expression
Mitochondrial inner membrane (see Table 3)
Sterol esterification for storage in lipid particles
Acyl-CoA:sterol acyltransferase (ACAT): ARE1[3,5]

Figure 5

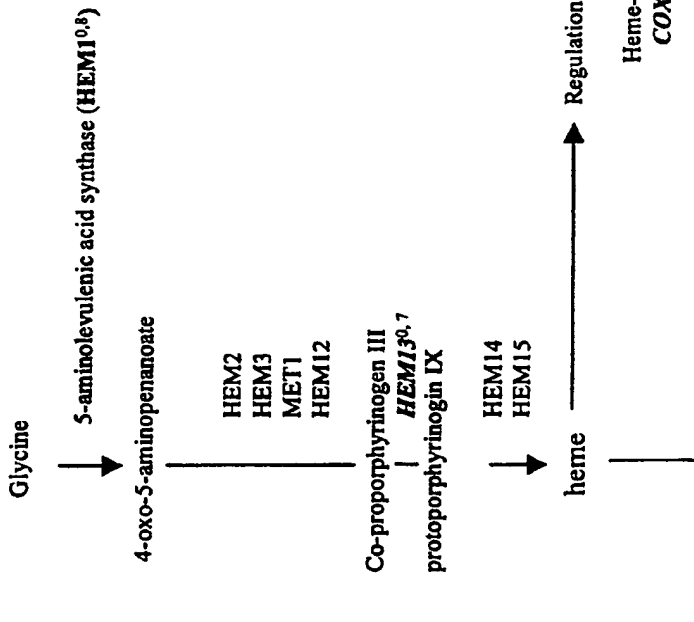

Glycine
5-aminolevulenic acid synthase (HEM1[0,8])
4-oxo-5-aminopenanoate
HEM2
HEM3
MET1
HEM12
Co-proporphyrinogen III
HEM13[0,7]
protoporphyrinogin IX
HEM14
HEM15
heme
Regulation of gene expression:
Heme-repressed:
COX5B[1,8], ERG11[3,3], NCP1[3,2], HMG2, OLE1[2,3], HEM13[0,7], ANB1
Heme-induced:
CYC1, CYT1[2,4], CYB2. COR1[2,3], COR2, COX4[3,4], COX5A[2,4],
COX6, CTT1, CTA1, SOD2[2,4], HMG1, TIF51A, ROX1
Heme-containing proteins:
COR1[2,3], COX1, 5A[2,4]/ 5B[1,8], CTT1,
CYB2, CYC1, CYC3, CYC7[1,5],
CYT1[2,4], CYT2, ERG11[3,3], 7, 9,
FRE1, HAP1, HEM1[0,8], HEM4,
HSP82, MET10, OLE1[2,7], PET100,
SHD3, 4, SOD2[2,4], YDJ1, YNL234W

Figure 6

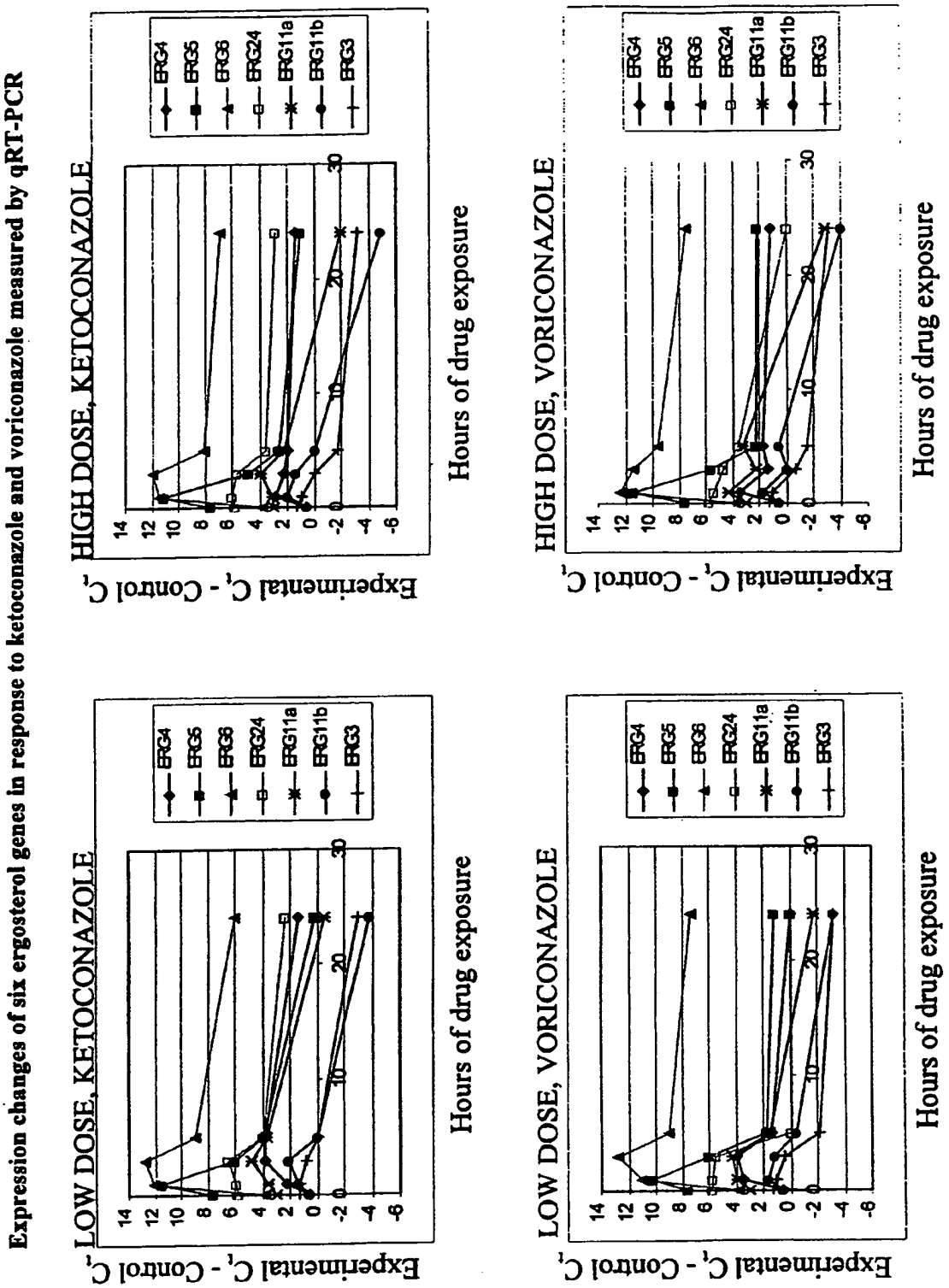

Expression changes of six ergosterol genes in response to ketoconazole and voriconazole measured by qRT-PCR
LOW DOSE, KETOCONAZOLE
HIGH DOSE, KETOCONAZOLE
LOW DOSE, VORICONAZOLE
HIGH DOSE, VORICONAZOLE
Experimental C_t - Control C_t
Hours of drug exposure
ERG4
ERG5
ERG6
ERG24
ERG11a
ERG11b
ERG3

Figure 7

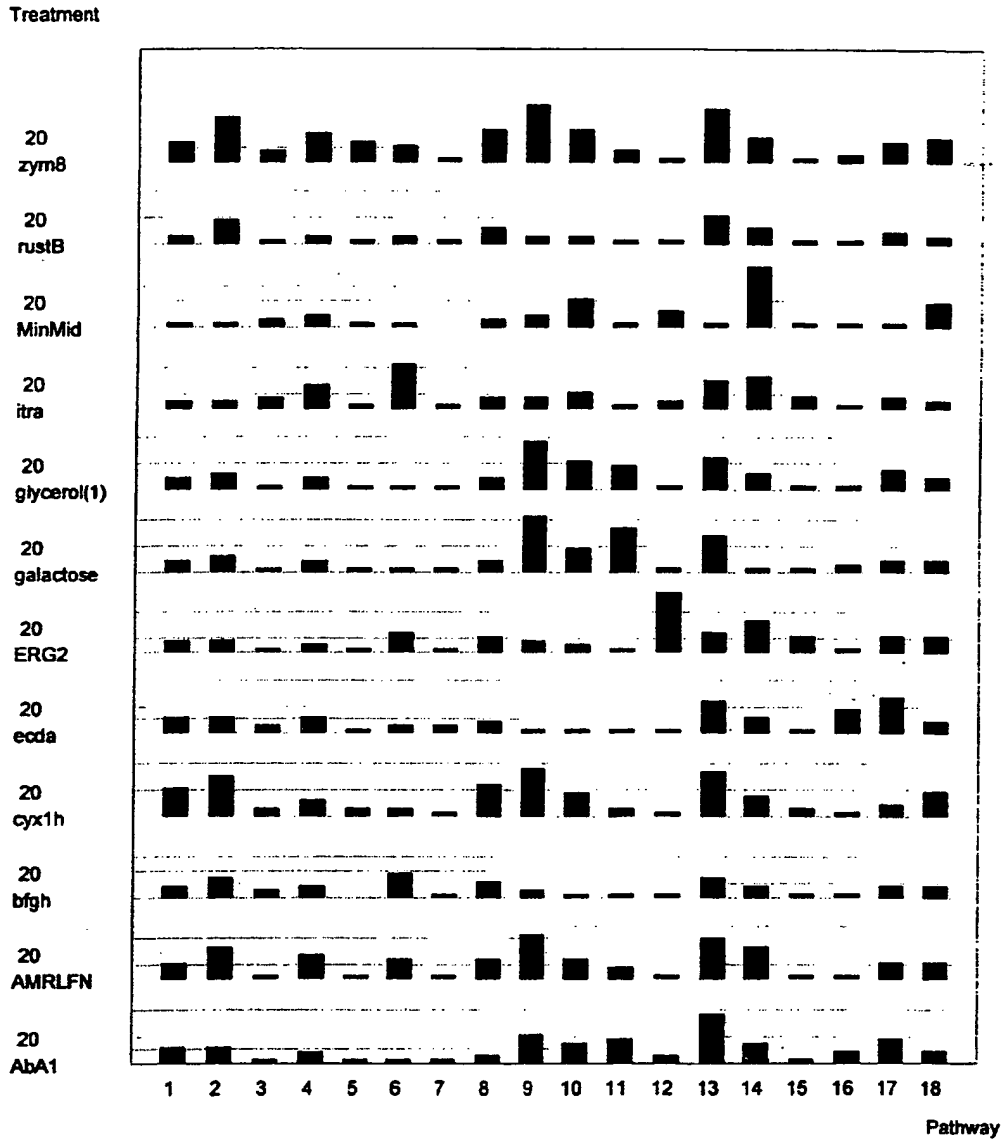

Treatment
20 zym8
20 rustB
20 MinMid
20 itra
20 glycerol(1)
20 galactose
20 ERG2
20 ecda
20 cyx1h
20 bfgh
20 AMRLFN
20 AbA1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Pathway

FIGURE 8

SEQUENCE LISTING

<110> Fostel, Jennifer M.
Bammert, Gary F.
Pharmacia Upjohn

<120> Gene Expression Profiles

<130> PUJ0290

<140>
<141>

<150> 60/156,050
<151> 1999-09-24

<150> 60/209,026
<151> 2000-06-02

<160> 21

<170> PatentIn Ver. 2.1

<210> 1
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 1
ccaaatggca atggaaatca a 21

<210> 2
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 2
acaaaccatg gaacggtcaa c 21

<210> 3

<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 3
ttggcagtca gtgctatccc atggatg 27

<210> 4
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 4
tcactttggg tgcctgtttg 20

<210> 5
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 5
acatgacaac ccccaattgt g 21

<210> 6
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 6
tgtcacatag ccgtaagtct cccactgct 29

<210> 7

<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 7
cgtgcatcaa cttacaccaa gc 22

<210> 8
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 8
gtgctcaaat ggctaaggat cat 23

<210> 9
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 9
taacaccaaa ctggcttacc acctgcagca a 31

<210> 10
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 10
agacacctac agaggaggaa tcaca 25

<210> 11

<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 11
gaccacagat ctctcaacta cttcaga 27

<210> 12
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 12
tgcaacccca actcccaaac ctttaagt 28

<210> 13
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 13
ccacctctag tgttttactg gattcc 26

<210> 14
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 14
cttcgaaaaa ctcgtatggc ttc 23

<210> 15

<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 15
taccgtacac aacagcacta ccgaccca 28

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 16
accacaggca atagacagtc cata 24

<210> 17
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 17
tgagctttgg aacggtatca ag 22

<210> 18
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 18
ttcacgattg cccagatagt gtagcgca 28

<210> 19

<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 19
gtagagttga aaccggtgtc atca 24

<210> 20
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 20
aacggacttg acttcagtgg taaca 25

<210> 21
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Novel Sequence

<400> 21
caggtatggt tgttactttt gccccagctg 30